



Memoria presentada para optar al título de Doctor por la
Universitat de Barcelona

Programa de Doctorat en Genètica, Bienni 2001-2003

**Polimorfismo y divergencia en genes
implicados en la resistencia al frío en
*Drosophila***

Tesi doctoral presentada per:
Carlos E. Arboleda Bustos

Directora de tesi:
Dra. Carmen Segarra Robert

Barcelona, junio 2008

Agradecimientos

Quiero expresar mi agradecimiento

A Carmen Segarra por su paciencia, apoyo, confianza y gran esfuerzo durante estos años.

A Montse Aguadé y a Julio Rozas por haberme brindado la oportunidad de realizar la tesis en el grupo y por su confianza.

A los parceros, la Zariguella y el Chapu, con ellos compartí grandes momentos inolvidables durante esta importante etapa de mi vida.

A Ursula por su gran amistad durante tantos años.

A Sebas y Filipe por su ayuda en la constante lucha con los diversos problemas informáticos.

A toda la gente del grupo de Genética Molecular Evolutiva.

A Montse Papaceit por haber realizado las hibridaciones *in situ*.

A Daniel Grinberg por haberme abierto las puertas del Departamento.

A Gabi y Mao por haber estado en las buenas y en las malas durante este tiempo.

A Sandruca por este último año lleno de bonitas sorpresas y alegrías.

Finalmente, este trabajo lo he llevado a cabo gracias a una beca de formación de personal investigador del Ministerio de Ciencia y Tecnología (FP 2001-0337). Además, el trabajo a contado con las subvenciones de la Secretaría General de Política Científica y Tecnológica (BMC2001-2906 y BFU2004-02253) y de la Generalitat de Catalunya como a un Grupo de Investigación Consolidado (2005SGR-166).

A mis padres
Pastora y Humberto

INTRODUCCIÓN	1
1. EVOLUCIÓN MOLECULAR	3
2. GENES IMPLICADOS EN LA ADAPTACIÓN AL FRÍO	9
El gen <i>Frost (Fst)</i>	12
El gen <i>Drosophila cold acclimation (Dca)</i>	13
El gen <i>Regucalcin (RC)</i>	14
El gen CG6296	15
3. <i>DROSOPHILA</i> COMO ORGANISMO MODELO	15
El género <i>Drosophila</i>	16
El grupo <i>melanogaster</i>	16
El grupo <i>obscura</i>	17
OBJETIVOS	19
MATERIALES Y MÉTODOS	23
1. MUESTRA POBLACIONAL	25
2. EXTRACCIÓN DE ADN	25
3. AMPLIFICACIÓN Y SECUENCIACIÓN	26
4. ENSAMBLAJE Y ALINEAMIENTO	31
5. HIBRIDACIÓN <i>IN SITU</i>	32
6. ANÁLISIS INTRAESPECÍFICOS E INTERESPECÍFICOS	32
7. ESTIMAS DE LA VARIABILIDAD NUCLEOTÍDICA INTRAESPECÍFICA	33
8. DIFERENCIACIÓN GENÉTICA	34
9. ESTIMAS DEL DESEQUILIBRIO DE LIGAMIENTO	35
10. ESTIMAS DE LA TASA DE RECOMBINACIÓN	35
11. TEST DE NEUTRALISMO	35
12. ANÁLISIS DE <i>MAXIMUM LIKELIHOOD (ML)</i>	37
RESULTADOS Y DISCUSIÓN	39
CAPÍTULO I: EL GEN FROST (<i>Fst</i>)	41
I.1. POLIMORFISMO NUCLEOTÍDICO DEL GEN <i>Fst</i>	43
I.1.1 RESULTADOS	43
I.1.1.1 POLIMORFISMO NUCLEOTÍDICO DEL GEN <i>Fst</i> EN <i>D. MELANOGASTER</i>	43
Diferenciación genética entre poblaciones	43

Análisis del polimorfismo nucleotídico	44
Análisis del polimorfismo aminoacídico	46
Diversidad haplotípica	48
Desequilibrio de ligamiento y recombinación	49
Diversidad Nucleotídica	49
Test del neutralismo	50
Genealogía de los alelos	53
I.1.1.2 POLIMORFISMO NUCLEOTÍDICO DEL GEN <i>Fst</i> EN <i>D. SUBOBSCURA</i>	55
Ubicación citológica del gen <i>Fst</i>	55
Análisis del polimorfismo nucleotídico	55
Análisis del polimorfismo aminoacídico	57
Diversidad haplotípica	58
Desequilibrio de ligamiento y recombinación	59
Diversidad nucleotídica	59
Test del neutralismo	60
Genealogía de los alelos	61
I.1.2 DISCUSIÓN	63
I.1.2.1 DIVERSIDAD NUCLEOTÍDICA DEL GEN <i>Fst</i> EN <i>D. MELANOGASTER</i> Y <i>D. SUBOBSCURA</i>	63
I.2 DIVERGENCIA DEL GEN <i>Fst</i>	71
I.2.1 RESULTADOS	71
I.2.1.1 DIVERGENCIA NUCLEOTÍDICA DEL GEN <i>Fst</i> EN EL GRUPO <i>MELANOGASTER</i> DE <i>DROSOPHILA</i>	71
I.2.1.1.1 DIVERGENCIA AMINOACÍDICA DE LA PROTEÍNA FROST EN EL GRUPO <i>MELANOGASTER</i> DE <i>DROSOPHILA</i>	74
I.2.1.2 DIVERGENCIA NUCLEOTÍDICA DEL GEN <i>Fst</i> EN EL GRUPO <i>OBSCURA</i> DE <i>DROSOPHILA</i>	77
I.2.1.2.1 DIVERGENCIA AMINOACÍDICA DE LA PROTEÍNA FROST EN EL GRUPO <i>OBSCURA</i> DE <i>DROSOPHILA</i>	79
I.2.2 DISCUSIÓN	81
I.2.2.1 DIVERGENCIA NUCLEOTÍDICA Y AMINOACÍDICA DEL GEN <i>Fst</i> EN <i>D. MELANOGASTER</i> Y <i>D. SUBOBSCURA</i>	81
CAPÍTULO II: EL GEN <i>DROSOPHILA COLD</i> <i>ACCLIMATION (Dca)</i>	87
II.1. POLIMORFISMO NUCLEOTÍDICO DEL GEN <i>Dca</i>	89

II.1.1 RESULTADOS	89
II.1.1.1 POLIMORFISMO NUCLEOTÍDICO DEL GEN <i>Dca</i> EN <i>D. MELANOGASTER</i>	89
Diferenciación genética entre poblaciones	89
Análisis del polimorfismo nucleotídico	90
Análisis del polimorfismo aminoacídico	92
Diversidad haplotípica	93
Desequilibrio de ligamiento y recombinación	93
Diversidad nucleotídica	94
Test del neutralismo	95
Genealogía de los alelos	96
II.1.1.2 POLIMORFISMO NUCLEOTÍDICO DEL GEN <i>DCA</i> EN <i>D. SUBOBSCURA</i>	97
Ubicación citológica del gen <i>Dca</i>	97
Análisis del polimorfismo nucleotídico	97
Análisis del polimorfismo aminoacídico	98
Diversidad haplotípica	100
Desequilibrio de ligamiento y recombinación	101
Diversidad nucleotídica	101
Test del neutralismo	102
Genealogía de los alelos	103
II.1.2 DISCUSIÓN	105
II.1.2.1 DIVERSIDAD NUCLEOTÍDICA DEL GEN <i>Dca</i> EN <i>D. MELANOGASTER</i> Y <i>D. SUBOBSCURA</i>	105
II.2 DIVERGENCIA DEL GEN <i>Dca</i>	109
II.2.1 RESULTADOS	109
II.2.1.1 DIVERGENCIA NUCLEOTÍDICA DEL GEN <i>Dca</i> EN <i>DROSOPHILA</i>	109
II.2.2 DISCUSIÓN	125
II.2.2.1 DIVERGENCIA NUCLEOTÍDICA DEL GEN <i>DCA</i> EN <i>DROSOPHILA</i>	125
Genes parálogos	127
CAPÍTULO III: EL GEN REGUCALCIN (<i>RC</i>)	129
III.1. POLIMORFISMO NUCLEOTÍDICO DEL GEN <i>RC</i>	131
III.1.1 RESULTADOS	131

III.1.1.1 POLIMORFISMO NUCLEOTÍDICO DEL GEN <i>RC</i> EN <i>D. MELANOGASTER</i>	131
Análisis del polimorfismo nucleotídico	131
Diversidad haplotípica	131
Desequilibrio de ligamiento y recombinación	134
Diversidad nucleotídica	134
Test del neutralismo	135
Genealogía de los alelos	136
III.1.1.2 POLIMORFISMO NUCLEOTÍDICO DEL GEN <i>RC</i> EN <i>D. SUBOBSCURA</i>	139
Ubicación citológica del gen <i>RC</i>	139
Análisis del polimorfismo nucleotídico	139
Desequilibrio de ligamiento y recombinación	140
Diversidad nucleotídica	141
Test del neutralismo	141
Genealogía de los alelos	143
III.1.2 DISCUSIÓN	145
III.1.2.1 DIVERSIDAD NUCLEOTÍDICA DEL GEN <i>RC</i> EN <i>D. MELANOGASTER</i> Y <i>D. SUBOBSCURA</i>	145
III.2 DIVERGENCIA DEL GEN <i>RC</i>	149
III.2.1 RESULTADOS	149
III.2.1.1 DIVERGENCIA NUCLEOTÍDICA DEL GEN <i>RC</i> EN <i>DROSOPHILA</i>	149
III.2.2 DISCUSIÓN	161
III.2.2.1 DIVERGENCIA NUCLEOTÍDICA DEL GEN <i>RC</i> EN <i>DROSOPHILA</i>	161
III. 3 LOS GENES PARÁLOGOS <i>Dca</i> y <i>RC</i>	165
III.3.1 RESULTADOS	165
III.3.1.1 EVOLUCIÓN MOLECULAR DE LOS GENES PARÁLOGOS <i>Dca</i> Y <i>RC</i> EN <i>D. MELANOGASTER</i>	165
III.3.1.2 EVOLUCIÓN MOLECULAR DE LOS GENES PARÁLOGOS <i>Dca</i> Y <i>RC</i> EN <i>D. SUBOBSCURA</i>	168
III.3.1.3 ANÁLISIS DE <i>MAXIMUM LIKELIHOOD (ML)</i> EN LA REGIÓN CODIFICADORA DE LOS GENES PARÁLOGOS <i>DCA</i> Y <i>RC (DR)</i>	171

III.3.2 DISCUSIÓN	175
III.3.2.1 EVOLUCIÓN MOLECULAR Y ANÁLISIS DE <i>ML</i> DE LOS GENES <i>DCA</i> Y <i>RC</i> EN <i>DROSOPHILA</i>	175
CAPÍTULO IV: EL GEN CG6296	179
IV.1 POLIMORFISMO NUCLEOTÍDICO DEL GEN CG6296	181
IV.1.1 RESULTADOS	181
IV.1.1.1 POLIMORFISMO NUCLEOTÍDICO DEL GEN CG6296 EN <i>D. MELANOGASTER</i>	181
Diferenciación genética entre poblaciones	181
Análisis del polimorfismo nucleotídico	182
Análisis del polimorfismo aminoacídico	184
Diversidad haplotípica	184
Desequilibrio de ligamiento y recombinación	185
Diversidad nucleotídica	185
Test del neutralismo	186
Genealogía de los alelos	187
IV.1.2 DISCUSIÓN	189
IV.1.2.1 DIVERSIDAD NUCLEOTÍDICA DEL GEN CG6296 EN <i>D. MELANOGASTER</i>	189
IV.2 DIVERGENCIA DEL GEN CG6296	191
IV.2.1 RESULTADOS	191
IV.2.1.1 DIVERGENCIA NUCLEOTÍDICA DEL GEN CG6296 EN EL GRUPO <i>MELANOGASTER</i> DE <i>DROSOPHILA</i>	191
IV.2.1.1.1 DIVERGENCIA AMINOACÍDICA DEL GEN CG6296 EN EL GRUPO <i>MELANOGASTER</i> DE <i>DROSOPHILA</i>	193
IV.2.2 DISCUSIÓN	197
IV.2.2.1 DIVERGENCIA NUCLEOTÍDICA Y AMINOACÍDICA DEL GEN CG6296 EN EL GRUPO <i>MELANOGASTER</i>	197
CONCLUSIONES	199
BIBLIOGRAFÍA	205

INTRODUCCIÓN

1. EVOLUCIÓN MOLECULAR

El estudio del origen y el mantenimiento de la variabilidad genética en las poblaciones naturales ha sido de gran importancia en los últimos 70 años. Este interés se debe al papel fundamental que la presencia de la variación genética tiene en el proceso evolutivo.

En el año 1968, Motoo Kimura propuso la teoría neutralista de la evolución molecular que sostiene que la mayor parte de los cambios evolutivos que se observan a nivel molecular se han producido por la fijación aleatoria de mutaciones selectivamente neutras. De hecho, de acuerdo al neutralismo la mayor parte de la variación molecular que se observa en las especies es selectivamente neutra y el resultado de un equilibrio entre la aparición de nuevas variantes por mutación y la eliminación o fijación de éstas por deriva genética. Así, desde el punto de vista de la teoría neutralista el polimorfismo es una fase transitoria de la evolución molecular. Además de las mutaciones neutras, la teoría neutralista también acepta la existencia de mutaciones deletéreas que son rápidamente eliminadas de la población por selección purificadora y, por lo tanto, no contribuyen ni al polimorfismo ni a la divergencia. De esta forma, la teoría neutralista considera que parte de las mutaciones presentan efectos deletéreos sobre la eficacia biológica, una gran fracción no afecta a la eficacia biológica y una porción muy pequeña y prácticamente despreciable confiere una ventaja selectiva a los individuos portadores.

Un objetivo importante en los estudios de genética de poblaciones a nivel molecular es detectar la huella dejada por la selección natural al actuar sobre las secuencias nucleotídicas. El nivel y el patrón de la variabilidad nucleotídica permiten detectar esta acción ya que ambos dependen no solamente de la mutación y la deriva genética sino también de la selección natural. Una nueva variante producida por mutación puede afectar la capacidad de supervivencia y reproducción del individuo portador. Se considera que una mutación es selectivamente ventajosa cuando incrementa la eficacia biológica y selectivamente deletérea cuando la disminuye. Por el contrario, las mutaciones neutras no tienen ningún efecto sobre la eficacia biológica. El destino de las mutaciones ventajosas y deletéreas depende fundamentalmente de la selección natural, mientras que el de las mutaciones neutras está regido únicamente por el azar. La mayor parte de las mutaciones deletéreas tienden a ser eliminadas, mientras

que las ventajosas pueden llegar a la fijación por selección positiva. La selección direccional positiva se produce cuando en una población surge por mutación una nueva variante ventajosa, es decir que confiere al individuo portador una mayor eficacia biológica con respecto a los demás. Esta nueva variante puede incrementar su frecuencia y llegar a la fijación en dicha población. La fijación de la variante seleccionada positivamente puede tener un efecto sobre las variantes neutras adyacentes tal como propusieron Maynard Smith y Haigh (1974). Dicho efecto se denomina *hitchhiking*, efecto de arrastre o barrido selectivo y consiste en la fijación junto con la variante seleccionada de las variantes neutras ligadas. El resultado es una disminución del nivel de variabilidad alrededor de la posición en que apareció la variante seleccionada. La región de ADN afectada por el barrido selectivo depende de la magnitud de la selección y la frecuencia de recombinación. Se ha observado que en regiones de baja recombinación este efecto es más acusado (Aguadé et al. 1989; Begun y Aquadro 1992; Berry et al. 1991; Martín-Campos et al. 1992), aunque también se ha detectado en regiones con niveles intermedios de recombinación (Aquadro et al. 1994; Aguadé y Langley 1994; Stephan 1994). Alternativamente, la selección puede conducir una nueva variante que ha aparecido por mutación no a la fijación sino a una frecuencia de equilibrio con el mantenimiento de dos o más variantes en la población (selección equilibradora o balanceadora). Los efectos de la selección balanceadora o equilibradora son opuestos a los de la selección direccional. En este caso, el mantenimiento de dos o más variantes en la población provoca que las variantes seleccionadas permanezcan en la población durante mucho más tiempo que el esperado si dichas variantes hubieran evolucionado únicamente por deriva genética. De esta forma se producirá una acumulación de variantes neutras en las posiciones estrechamente ligadas a la posición sometida a selección. El tamaño de la región afectada por el exceso de variación también depende de la frecuencia de recombinación.

A pesar de que es tentador atribuir la desviación del modelo neutro a la acción de la selección positiva, la historia demográfica (la trayectoria de las especies a través del tiempo) puede tener efectos similares a los de la selección en los patrones de variabilidad. Por ejemplo, una fuerte reducción en los niveles de variación nucleotídica puede ser el resultado de un arrastre selectivo (*selective sweep*) o de un fuerte cuello de botella sufrido por la especie. Para distinguir efectos selectivos y demográficos es necesario estudiar múltiples regiones genómicas (*multilocus approach*). El razonamiento tras esta aproximación es que mientras la historia demográfica afecta los

patrones de variabilidad del genoma en su totalidad, la acción de la selección positiva actúa localmente en genes individuales. El poder cuantificar la importancia relativa que tienen la selección natural y los factores demográficos en moldear la variabilidad nucleotídica es actualmente uno de los temas de más debate en la genética de poblaciones molecular. La aproximación genómica para hacerla se ha aplicado especialmente en poblaciones africanas y no africanas de *Drosophila melanogaster* (Begun y Whitley 2000; Andolfatto 2001; Kauer et al. 2002; Wall et al. 2002; Glinka et al. 2003; Orengo y Agudé 2004). Varios de los trabajos recientes han concluido que la desviación del modelo neutro observado en las poblaciones no africanas (Harr et al. 2002; Glinka et al. 2003; Kauer et al. 2003; Orengo y Agudé 2004) y algunas poblaciones africanas (Andolfatto y Przeworski 2001; Mousset et al. 2003) son mejor explicados por la acción de la selección natural que por la historia demográfica de la especie.

Por otro lado, los cambios adaptativos a nivel proteico producidos en un determinado linaje pueden detectarse por un incremento en la tasa de fijación de mutaciones no sinónimas en dicho linaje. Este tipo de aproximación ha revelado que la evolución por selección positiva es más frecuente de lo que se pensaba anteriormente, por lo menos en *Drosophila* (Smith y Eyre-Walker 2002).

La rápida acumulación de datos de secuencias de ADN durante los últimos años, ha proporcionado evidencia tanto de patrones neutros como de patrones selectivos o adaptativos. La comparación entre la variación genética entre especies cercanas (divergencia) y la variación genética en los individuos de una misma especie (polimorfismo) es una poderosa vía para estudiar y determinar las fuerzas que están causando la evolución molecular.

Otro mecanismo de evolución molecular es la duplicación génica. Recientemente, la secuenciación del genoma de diferentes organismos modelo han permitido determinar que este mecanismo es importante para explicar el origen de nuevos genes y de nuevos sistemas genéticos. La importancia de la duplicación génica en la creación de nuevos genes fue apuntada por primera vez por Bridges (1935) y Muller (1936). Ambos autores mostraron que los cromosomas politénicos de *Drosophila melanogaster* contenían abundantes duplicaciones intracromosómicas pequeñas y que algunas mutaciones, como la que causa el fenotipo *Bar*, eran provocadas por duplicaciones. Este hallazgo fue el punto de partida para proponer que la duplicación génica es un importante mecanismo de origen de nuevos genes (Lewis

1951; Stephens 1951). Sin embargo, las evidencias moleculares de esta idea fueron obtenidas por Ingram (1961, 1963) al demostrar que las cadenas γ , α , y β de la hemoglobina y la mioglobina de los humanos son los productos de una familia de genes que surgieron por sucesivas duplicaciones que ocurrieron hace mucho tiempo. En los años 60 se descubrieron diferentes familias multigénicas, lo que permitió estudiar su evolución. Sin embargo, la magnitud real de la importancia de la duplicación génica no fue reconocida hasta que fueron disponibles las secuencias genómicas de diferentes organismos modelo.

Existen diversos mecanismos por los que puede producirse la duplicación de genes: el entrecruzamiento desigual, la retroposición y la duplicación cromosómica. Sin embargo, el resultado de estos mecanismos puede ser diferente. El entrecruzamiento desigual produce una serie de duplicaciones génicas que se ubican de forma consecutiva, es decir, en tándem en el cromosoma. Dependiendo de la posición del entrecruzamiento, la región duplicada puede contener parte de un gen, el gen completo o varios genes. En los últimos dos casos, si los genes originales presentan intrones, éstos también estarán presentes en los genes duplicados. Además, el entrecruzamiento desigual ha sido importante para la formación de nuevos genes a partir de la combinación de exones presentes en distintos genes (*exon shuffling*). Por otro lado, la retroposición ocurre cuando un RNA mensajero se retrotranscribe a ADN y posteriormente se inserta en el genoma. Los genes duplicados resultantes (retrogenes) carecen de las secuencias reguladoras y de todos o parte de los intrones y pueden presentar secuencias de poliadenilación y en algunos casos repeticiones de regiones cortas. Otra de las diferencias con el cruzamiento desigual, es que un gen originado por retroposición normalmente no se encuentra ligado al gen original debido a que la inserción en el genoma se produce aleatoriamente. Puesto que el promotor y las secuencias reguladoras del gen no se transcriben y por lo tanto no están presentes en la copia originada por retroposición, el gen duplicado resultante normalmente carece de los elementos necesarios para la transcripción y se convierte en un pseudogen. Sin embargo, se han descrito diversos casos de genes duplicados por retroposición que se expresan, probablemente debido a que se insertan en una región genómica que se encuentra localizada *downstream* de una región promotora (Long et al. 2003). Finalmente, la duplicación cromosómica o de genomas completos puede originarse por una mitosis abortiva en la que se produce la duplicación cromosómica pero no la

división celular. Existen evidencias de que este tipo de duplicación a gran escala ocurre frecuentemente en plantas pero no en animales (Li 1997).

Similarmente a las mutaciones puntuales, las duplicaciones génicas pueden llegar a fijarse o a perderse en la población. La probabilidad de fijación de los genes duplicados selectivamente neutros es muy pequeña ($1/2N$) lo que sugiere que muchos de ellos son eliminados de la población. En promedio, el tiempo que se requiere para llegar a su fijación es de $4N$ generaciones (Kimura 1983).

En muchos casos tener dos copias idénticas de un gen no es desventajoso puesto que genera redundancia funcional. Por lo tanto, las mutaciones que producen modificaciones en la estructura y la función en uno de los dos genes no son eliminadas por selección ya que no son deletéreas. Con el tiempo, el acumulo de este tipo de mutaciones produce pérdida de expresión y/o de función del gen convirtiéndolo en un pseudogen. Finalmente, el pseudogen no será identificable debido a que es eliminado del genoma o a la elevada divergencia entre el pseudogen y la copia funcional. El proceso de pseudogenización usualmente ocurre durante los primeros millones de años después de la duplicación si el gen duplicado no se encuentra bajo ningún tipo de selección (Lynch y Conery 2000). Sin embargo, antes de convertirse en un pseudogen debido a la relajación de la restricción funcional, algunos genes duplicados pueden permanecer un largo tiempo en el genoma realizando una función específica. Por ejemplo, en humanos y ratones el número de genes implicados en el sistema olfativo es muy similar, pero el porcentaje de pseudogenes en los humanos (>60%) es mucho mayor que en los ratones (20%). Además, los pseudogenes ocasionalmente pueden servir para realizar alguna función o ser reactivados (Otha 1994; Trabesinger-Ruef et al. 1996; Kleineidam et al. 1999).

Por otro lado, algunas veces la presencia de genes duplicados es beneficiosa debido a la cantidad extra de proteína o RNA que se produce. El modo de evolución que permite que dos genes parálogos puedan tener una secuencia y función similares se denomina evolución concertada (Li 1997). Uno de los mecanismos que explica el mantenimiento de la misma función de los genes parálogos después de la duplicación es la conversión génica. Otra alternativa para prevenir la diferenciación entre los genes es la selección purificadora en contra de las mutaciones que producen cambios en el funcionamiento del gen. Se ha observado en varias familias de genes que se piensa evolucionan de acuerdo al modelo de evolución concertada que la selección

purificadora es más importante que la conversión génica en mantener funciones comunes de los genes duplicados (Nei et al. 2000; Piontkivska et al. 2002).

Cuando los genes duplicados difieren en algún aspecto de su función se pueden mantener por subfuncionalización (Cooke et al. 1997). En este caso, después de la duplicación cada gen adopta parte de la función original del gen ancestral. Una forma de subfuncionalización importante es la división de la expresión génica (Force et al. 1999). Los cambios en la expresión génica después de la duplicación parece ser una regla general más que una excepción y normalmente ocurren rápidamente después de la duplicación génica (Wagner 2000; Gu et al. 2002). Además, a nivel del funcionamiento de la proteína la subfuncionalización puede llevar a una especialización funcional cuando uno de los genes duplicados adquiere la capacidad de llevar a cabo mejor una de las funciones originales del gen ancestral (Hughes 1999). Sin embargo, no está claro que porcentaje de genes duplicados evolucionan por subfuncionalización.

Finalmente, la neofuncionalización o el origen de una función nueva es uno de los resultados más importantes de la duplicación génica. La complejidad de los genomas podría haber aumentado como consecuencia del proceso de duplicación y posterior diferenciación de las copias lo que produciría nuevos genes con funciones nuevas. La relajación de la restricción selectiva puede ser importante en la adquisición de nuevas funciones. Mientras que una de las copias de un gen duplicado mantiene la función, la otra puede acumular cambios que con el tiempo lo llevarían a adquirir una función nueva. Por otro lado, la acción de la selección positiva a favor de los cambios ventajosos podría incrementar el número de estas sustituciones y producir una nueva función. Además, la selección positiva debe jugar un papel fundamental para que la nueva copia se mantenga y no sea convertida en un gen no funcional. Por lo tanto, este tipo de selección sería la explicación más plausible para la adquisición de nuevas funciones (Otha 1994).

Una aproximación para detectar la acción de la selección natural positiva es el estudio del polimorfismo y la divergencia en genes candidatos de haber sufrido cambios adaptativos, bien sea en genes de copia única o pertenecientes a familias multigénicas. Esta aproximación ha dado resultados positivos en genes que están involucrados principalmente en la reproducción (Civetta y Singh 1995, Tsaur y Wu 1997, Aguadé 1999, Begun et al. 2000, Swanson et al. 2001) o en la interacción entre huésped-patógeno (Yang y Bielawski 2000), entre otros.

Los genes implicados en la tolerancia, aclimatación y respuesta al frío son buenos candidatos de haber sufrido cambios adaptativos. De hecho, es incuestionable el carácter adaptativo de la aclimatación de las especies a las condiciones ambientales. La capacidad de las especies de colonizar nuevas áreas dependerá de su capacidad de adaptación a las nuevas condiciones ambientales. Así, por ejemplo, la posible expansión de las especies a regiones más frías debería ir acompañada de cambios moleculares que confieran una mayor capacidad de supervivencia a bajas temperaturas en los genes implicados en la respuesta al frío.

En el presente trabajo se pretende analizar la dinámica evolutiva de genes implicados en la tolerancia, aclimatación y respuesta al frío a partir del estudio de los patrones de polimorfismo intraespecífico en *D. melanogaster* y *D. subobscura* y de la divergencia interespecífica en varias especies del grupo *melanogaster* y del grupo *obscura* de *Drosophila*.

2. GENES IMPLICADOS EN LA ADAPTACIÓN AL FRÍO

La temperatura es uno de los factores medioambientales más importantes que influyen en la biología de los animales debido a que afecta directamente a los procesos bioquímicos fundamentales que constituyen la vida. Sin embargo, todos los organismos han desarrollado un amplio rango de respuestas adaptativas que los protegen de los efectos perjudiciales causados por la variación de la temperatura (Cossins et al. 1987). Los periodos de baja temperatura son ubicuos en una gran parte de nuestro planeta, y tanto los animales como las plantas han adquirido diversas adaptaciones para sobrevivir a las bajas temperaturas. La mayoría de los organismos buscan algún refugio para protegerse de las bajas temperaturas como permanecer bajo el agua, la tierra o la nieve durante el invierno.

Para poder impedir la congelación de los líquidos corporales a temperaturas por debajo del punto de congelación, que se sitúa entre -0.5°C en las especies terrestres y -1.8°C en invertebrados marinos, los animales ectodermos utilizan dos estrategias principales: tolerar la congelación o impedirla (Storey y Storey 1986). Tolerar la congelación es la habilidad de poder regular y soportar la formación de hielo extracelular mientras se mantiene el estado líquido del citoplasma. Esta estrategia es característica de muchas especies de insectos terrestres (especialmente Himenópteros, Dípteros, Coleópteros y Lepidópteros), varios invertebrados marinos y los reptiles. La

mayoría de especies que toleran la congelación son capaces de convertir el 50-60% del total de agua corporal en hielo extracelular y han desarrollado una serie de mecanismos de defensa que las protegen de las consecuencias producidas por la formación interna de hielo (Storey 1999). Este tipo de mecanismos se basan principalmente en la producción de núcleos de hielo (generalmente proteínas plasmáticas) que permiten el inicio de la congelación a temperaturas bajo cero y en sustancias crioprotectoras que contribuyen a controlar la propagación del hielo dentro del cuerpo (Storey y Storey 1988).

Sin embargo, la gran mayoría de los organismos que viven en ambientes con periodos fríos optan por impedir la congelación. Para algunos es suficiente con encontrar un sitio de hibernación bien protegido de las bajas temperaturas y principalmente de la congelación. Otros, suprimen la congelación de los fluidos corporales empleando varias adaptaciones bioquímicas y fisiológicas que permiten mantener su estado líquido a temperaturas muy inferiores a los 0°C. Estas adaptaciones comprenden: la eliminación de proteínas nucleadoras de hielo, la acumulación de azúcares y polyols (por ejemplo, la trehalosa o el glicerol) que disminuyen la temperatura de cristalización y estabilizan las membranas a temperaturas bajas, y la síntesis de proteínas anticongelantes que reducen la posible formación de núcleos de hielo a partir de pequeños cristales (Sinclair et al. 2003).

Las estrategias tanto de tolerar la congelación como de impedirla han aparecido varias veces independientemente en grupos de animales que no se encuentran relacionados filogenéticamente. El uso de una determinada estrategia en una especie probablemente se deba a la interacción de varios factores como son: la necesidad de mantener la movilidad y un estilo de vida activo a bajas temperaturas, las condiciones ambientales del hábitat o lugar de hibernación, las capacidades fisiológicas pre-existentes de la especie, y la habilidad de desarrollar adaptaciones metabólicas y fisiológicas para perfeccionar una de las estrategias de tolerancia al frío (Storey y Storey 1986).

Las especies del género *Drosophila* son un buen modelo para realizar estudios de respuesta al estrés en animales ectodermos debido a su amplio rango de distribución latitudinal. En particular, *D. melanogaster* es una especie que ha podido adaptarse con éxito a nuevas condiciones ambientales y así colonizar nuevas áreas hasta alcanzar su actual distribución cosmopolita (David y Capy 1988). Las diferentes poblaciones de esta especie presentan adaptaciones locales que determinan patrones fenotípicos y de

variación genética complejos dentro y entre poblaciones (Hoffmann et al. 2003). Estudios realizados durante los últimos años con *D. melanogaster* han demostrado la presencia de clinas latitudinales para varios caracteres morfológicos (Capy et al. 1993), así como para caracteres directamente relacionados con la eficacia biológica (*fitness*) tales como la viabilidad, tasa de desarrollo, habilidad competitiva de las larvas, producción de huevos y tolerancia al estrés causado principalmente por el calor, el frío, la desecación y la inanición (Guerra et al. 1997; Karan et al. 1998; Hoffmann et al. 2003). En estudios centrados en fenotipos de tolerancia a la temperatura realizados en el laboratorio, se ha identificado una gran varianza genética en la resistencia a las altas y bajas temperaturas (Cavicchi et al. 1995; Loeschke y Krebs 1996; Bubli et al. 1998; Norry et al. 2004; Anderson et al. 2005). Además, se ha cuantificado la variación genética a lo largo de las clinas de resistencia al calor y al frío (Gilbert y Huey 2001; Hoffmann et al. 2002; Ayrinhac et al. 2004; Kimura 2004). Asimismo, se han detectado asociaciones de la variación fenotípica de la tolerancia a la temperatura con el polimorfismo nucleotídico y con la expresión de genes candidatos que afectan la resistencia al estrés térmico (McColl et al. 1996; Krebs y Feder 1997; Anderson et al. 2003).

A pesar de las evidencias que existen sobre el alto grado de variación genética que presentan los fenotipos de resistencia al estrés causado por el calor y el frío en poblaciones de *D. melanogaster*, es fundamental poder identificar genes que estén involucrados directamente con la variación observada para poder entender los mecanismos genéticos y evolutivos que influyen en la adaptación al estrés térmico, especialmente el causado por el frío.

Los estudios realizados para determinar la base genética de la tolerancia a la temperatura se han enfocado principalmente en identificar genes involucrados en la respuesta al estrés producido por calor (*Heat shock proteins, Hsps*). Así por ejemplo, los genes *Hsp70* (Feder y Krebs 1997; Krebs y Feder 1998; Krebs y Feder 1997; Feder et al. 1996; Krebs 1999; Dahlgaard et al. 1998; Bettencourt et al. 2002), *Hsp68* (McColl et al. 1996), *Hsr-omega* (McColl et al. 1996; McColl y McKechnie 1999; Anderson et al. 2003) y el factor *heat-shock (Hsf)* (Lerman y Feder 2001) han sido identificados como genes implicados en la resistencia al estrés producido por calor. Estos estudios se han realizado mediante el análisis de la variación clinal de las frecuencias alélicas, el análisis del cambio en la expresión génica en poblaciones con diferentes perfiles de resistencia al estrés causado por el calor, o a través del análisis del cambio sobre los

fenotipos resistentes causado al incrementar el número de copias del gen en cuestión. Sin embargo, es mucho menor el número de estudios realizados sobre la resistencia del estrés causado por el frío, posiblemente debido a la poca información sobre genes implicados en esta característica y/o a la falta de conocimiento sobre la vía metabólica afectada por este tipo de estrés. No obstante, Goto (2000, 2001) observó que en *Drosophila* los genes *Frost* (*Fst*) y *Drosophila cold acclimation gene* (*Dca*) se ven altamente sobreexpresados en respuesta a la aclimatización al frío.

En el presente trabajo se pretende analizar la dinámica evolutiva de genes implicados en la tolerancia, aclimatación y respuesta al frío a partir de los patrones de polimorfismo intraespecífico en *D. melanogaster* y *D. subobscura* y de la divergencia interespecífica en especies del grupo *melanogaster* y del grupo *obscura* de *Drosophila*. Los genes que se han elegido como candidatos de haber sufrido cambios adaptativos de tolerancia, aclimatación o resistencia al frío fueron en un principio *Fst* y *Dca*. Ambos genes fueron escogidos debido a los datos experimentales que mostraban su directa implicación en la respuesta al frío en *D. melanogaster*. Posteriormente, el estudio se amplió a los genes *Regucalcin* (*RC*) y *CG6296*, aunque la acción de estos genes no se ha relacionado con la resistencia al frío. La proteína codificada por el gen *RC* presenta una alta similitud con la secuencia aminoacídica de la proteína codificada por el gen *Dca*, lo que indica que ambos genes han surgido tras un evento de duplicación génica. Por último, la proteína codificada por el gen *CG6296* presenta una región de motivos PEEST similar a la detectada en la proteína producida por el gen *Fst*.

El gen *Frost* (*Fst*)

El gen *Fst* está altamente sobreexpresado durante la fase de recuperación a 25°C en individuos de *D. melanogaster* que han sido expuestos a un choque térmico a 0°C (Goto 2001). Además, la prolongación del choque frío es más eficiente en la inducción de la expresión del gen (Goto 2001; Qin et al. 2005; Morgan y Mackay 2006, Sinclair et al. 2007).

A pesar de que la función exacta del gen *Fst* no ha sido caracterizada, la proteína codificada por dicho gen presenta similitud con las proteínas de la familia MUCIN (Morgan y Mackay 2006). El gen *Fst* se encuentra ubicado en la banda 85E del brazo cromosómico 3R de *D. melanogaster*. La secuencia nucleotídica completa de la región

codificadora de este gen que no contiene ningún intrón está constituida por 834 pb y codifica una proteína de 278 aminoácidos. La secuencia aminoacídica de esta proteína presenta unos motivos PEEST (Pro-Glu-Glu-Ser-Thr) que se encuentran repetidos varias veces en la región C-terminal y una zona rica en prolinas en el centro de la proteína.

El perfil de hidrofobicidad sugiere que la proteína *Fst* tiene una estructura modular con una pequeña región hidrofóbica en la región N-terminal seguida de una región hidrofílica. Los 18 primeros aminoácidos de la región N-terminal presentan las características de un péptido señal, lo que sugiere que esta proteína es dirigida hacia el retículo endoplasmático (RE) y probablemente entre en la vía secretora. Sin embargo, en la región C-terminal no hay indicios de secuencias aminoacídicas particulares que puedan indicar que esta proteína permanezca en el RE. Al parecer la proteína *Fst* es excretada en el espacio extracelular.

El gen *Drosophila cold acclimation* (*Dca*)

El gen *Dca* fue identificado en un estudio de expresión génica durante la aclimatización a una temperatura moderadamente baja en *D. melanogaster* (Goto 2000). Este gen se ve altamente sobreexpresado durante la fase de recuperación a 15°C de individuos que han sido expuestos a un choque térmico a 0°C (Goto 2000; Qin et al. 2005). Además, el gen *Dca* presenta unos niveles de expresión que son superiores en individuos que han completado su desarrollo embrionario a 15°C desde la fase de huevo respecto a aquellos que lo han realizado a 25°C. El gen *Dca* de *D. melanogaster* presenta una elevada similitud, tanto a nivel de secuencia de nucleótidos (~55%) como de secuencia de aminoácidos de la proteína codificada (~34.8%), con el gen *senescence marker protein-30* (*smp-30*) clonado y secuenciado en rata, ratón y humano (Yamaguchi y Yamamoto 1978; Shimokawa y Yamaguchi 1993). Por lo tanto, *Dca* es considerado como el gen de *Drosophila* ortólogo del gen *smp-30* presente en mamíferos (Goto 2000).

El gen *Dca* de *D. melanogaster* se encuentra localizado en la banda 88D del brazo derecho del cromosoma 3. Este gen está formado por dos exones separados por un pequeño intrón de aproximadamente 43 pb. El gen *Dca* presenta una pauta de lectura de 909 pb y codifica para una proteína de 303 aminoácidos. En la región 5' flanqueante

se ha identificado un sitio de unión a AP-1 (un factor de la transcripción de *smp-30*) así como un elemento STRE (elemento de respuesta al estrés) y diferentes sitios de unión de otros factores de la transcripción.

Los estudios de expresión realizados en mamíferos del gen *smp-30* han revelado que dicho gen presenta un nivel de expresión alto durante todo el desarrollo (Fujita et al. 1996) pero que disminuye en la etapa de senescencia en ambos sexos (Fujita y Maruyama 1991; Fujita et al. 1992a,1992b). La función de la proteína SMP-30 se ha relacionado con el transporte de Ca^{2+} (Shimokawa y Yamaguchi 1993) y al parecer activa la proteína $(\text{Ca}^{2+}\text{-Mg}^{2+})\text{-ATPasa}$ que participa en la eliminación de Ca^{2+} del citosol a través de la membrana plasmática (Takahashi y Yamaguchi 1993a, 1993b, 1994; Fujita et al. 1998). Además, como la estructura primaria de la proteína codificada por el gen *smp-30* no presenta motivos conocidos de unión al Ca^{2+} (como los *EF-hans motif*), se ha sugerido que SMP-30 representa una nueva clase de proteína de unión al Ca^{2+} (Shimokawa y Yamaguchi 1993; Fujita et al. 1995; Fujita 1999; Yamaguchi 2005). Por lo tanto, es posible que el gen *Dca* también contribuya de manera directa en el mantenimiento de los niveles de Ca^{2+} en el citosol.

A diferencia del gen *smp-30* de los mamíferos, la expresión del gen *Dca* en *Drosophila* aumenta en la etapa de senescencia. Shi et al (1994) observaron que los niveles de absorción de ATP dependiente de Ca^{2+} aumentan en individuos adultos de *Drosophila*, lo que sugiere una conexión entre la sobreexpresión del gen *Dca* y los niveles de Ca^{2+} en esta etapa. Posteriormente, Fujita et al (1998) sugirieron que el gen *smp-30* regula la actividad de bombeo del Ca^{2+} y por lo tanto contribuye al rescate de la apoptosis. Además, Perotti et al (1990) mostraron que después de someter células humanas a un choque térmico a bajas temperaturas se producía la activación de endonucleasas y de la apoptosis debido a una prolongada elevación en los niveles de Ca^{2+} citosólico. Por consiguiente, puede ser que el incremento en los niveles de expresión del gen *Dca* pueda relacionarse con el mantenimiento de los niveles de Ca^{2+} a bajas temperaturas y el aumento de la tolerancia al frío.

El gen *Regucalin* (RC)

El gen *RC* de *Drosophila* presenta una elevada similitud nucleotídica con la del gen *Dca*, lo que indica que ambos genes surgieron tras un proceso de duplicación génica. Las proteínas codificadas por ambos genes también presentan una elevada

similitud y pertenecen a la familia de proteínas SMP-30. Por lo tanto, la función del gen *RC* podría estar también relacionada con el mantenimiento de los niveles Ca^{2+} , aunque no se ha demostrado su relación con la resistencia al frío. El gen *Regucalcin* está compuesto por dos exones y un pequeño intrón de aproximadamente 56 pb. Este gen presenta una pauta de lectura de 909 pb y codifica una proteína de 303 aminoácidos. El gen *RC* está localizado en la banda *I1A* del cromosoma *X* de *D. melanogaster*.

El gen CG6296

El gen CG6296 en *D. melanogaster* está compuesto por tres exones separados por dos intrones de aproximadamente 54 pb. Este gen presenta una pauta de lectura de 2022 pb y codifica una proteína de 674 aminoácidos. Al igual que el gen *Fst*, la secuencia aminoacídica de la proteína codificada por el gen CG6296 presenta unos motivos PEEST que se repiten varias veces en la región C-terminal de la proteína. La presencia de esta característica común con el gen *Fst* fue lo que motivo su inclusión en este estudio aunque CG6296 no se había relacionado con la resistencia al frío. Este gen se encuentra localizado en la banda *97D* del brazo cromosómico *3R* de *D. melanogaster*. La posible función descrita para la proteína CG6296 es de triacilglicerol lipasa, implicada en el transporte y el metabolismo de lípidos.

3. DROSOPHILA COMO ORGANISMO MODELO

Las características que presentan las especies del género *Drosophila* las han convertido en un sistema biológico modelo. La facilidad de captura y mantenimiento, el corto tiempo de generación y el elevado número de descendientes han permitido realizar una gran cantidad de estudios utilizando especies de este género. Dichos estudios abarcan campos tan dispares como la biología de poblaciones, sistemática, ecología, biología molecular y genética, entre otros.

En general, los drosofilidos son dípteros que se han especializado en obtener el alimento a partir de productos vegetales en descomposición, ya que su fuente principal de alimentación son las levaduras y las bacterias que fermentan estos materiales. La filogenia del género *Drosophila* se ha inferido empleando diferentes tipos de datos. Así, se han realizado filogenias a partir de datos biogeográficos, del registro fósil y morfológicos, a partir de la observación y comparación de los patrones de bandas de los

cromosomas politénicos, gracias a estudios de hibridación de ADN y, más recientemente, mediante la comparación de secuencias de ADN de genes mitocondriales o nucleares. La combinación de todos estos datos ha mejorado sustancialmente la comprensión de la filogenia del género.

El género *Drosophila*

La familia *Drosophilidae* incluye más de 3000 especies, de las que alrededor de unas 2000 pertenecen al género *Drosophila*. Este género se divide en tres subgéneros: *Sophophora*, *Drosophila* y *Idiomya*. En el presente trabajo se estudian algunas de las especies pertenecientes al subgénero *Sophophora* que está compuesto por cuatro grandes grupos: *melanogaster*, *obscura*, *willistoni* y *saltans*. Los estudios realizados durante los últimos años han permitido determinar relativamente bien las relaciones filogenéticas del subgénero (O'Grady y Kidwell 2002).

El grupo *melanogaster*

El grupo *melanogaster* de *Drosophila* está compuesto por aproximadamente unas 200 especies clasificadas en 10 subgrupos (Powell 1997). La distribución de la mayoría de los subgrupos sugiere que este grupo de especies se ha originado en África y posteriormente expandido por los cinco continentes principalmente gracias a la actividad humana. Las especies del subgrupo *melanogaster* y *ananassae* han sido, con diferencia, las más estudiadas del grupo. Las filogenias reconstruidas de acuerdo a diferentes métodos han mostrado generalmente la misma topología para los diferentes subgrupos. Sin embargo, el aporte de nuevos datos moleculares en los últimos años ha modificado las relaciones inferidas para algunas de las especies del subgrupo (Ko et al. 2003). El subgrupo *melanogaster* está formado por 9 especies: *D. melanogaster*, *D. simulans*, *D. sechellia*, *D. mauritiana*, *D. teissieri*, *D. yakuba*, *D. erecta*, *D. orena* y *D. santomea*. La distribución de *D. melanogaster* y *D. simulans* es cosmopolita, mientras que las otras siete especies se encuentran en la región afrotropical. *D. sechellia*, *D. mauritiana* y *D. santomea* son endémicas de las islas Seychellia, la isla Mauricio y la isla de São Tomé, respectivamente, mientras que el resto de especies están distribuidas por el continente.

El grupo *obscura*

El grupo *obscura* de *Drosophila* se originó en África y posteriormente colonizó las regiones templadas del norte, tanto de la zona Paleártica como Neártica, donde actualmente se encuentran un gran número de representantes. Recientes datos moleculares (Goddard et al. 1990) indican que la separación entre las especies paleárticas y neárticas es más reciente de lo que anteriormente se consideraba y que posiblemente las especies de América del Norte se originaron tras la migración de una población del Viejo Mundo a través del estrecho de Bering. El grupo *obscura* está formado por aproximadamente unas 35 especies y se divide en cuatro subgrupos: *obscura*, *pseudoobscura*, *affinis* y *microlabis*. El subgrupo *obscura* lo integran las especies del Viejo Continente, mientras que los subgrupos *pseudoobscura* y *affinis* reúnen las especies del Nuevo Mundo. El subgrupo *microlabis* está formado por especies africanas.

OBJETIVOS

Los estudios de polimorfismo nucleotídico intraespecífico y de divergencia interespecífica son una aproximación muy potente para detectar a nivel molecular la huella dejada por la acción de la selección natural adaptativa. El objetivo global del presente proyecto es aplicar esta aproximación en genes candidatos a haber sufrido cambios adaptativos y más concretamente en genes implicados en la resistencia al frío y en genes relacionados.

Este objetivo global se puede desglosar en los siguientes objetivos específicos:

1- Estudiar el nivel y el patrón del polimorfismo nucleotídico de los genes *Fst* y *Dca* en poblaciones naturales de *D. melanogaster* con el objetivo de detectar la huella de la acción de la selección natural al actuar sobre estos genes candidatos a haber sufrido cambios adaptativos al estar implicados en la resistencia y tolerancia al frío.

2- Comparar el nivel y el patrón del polimorfismo nucleotídico de los genes *Fst* y *Dca* (y también del gen *CG6296*) en dos poblaciones naturales de *D. melanogaster* ubicadas al sur (Montemayor, Córdoba) y al norte (Sant Sadurní d'Anoia, Barcelona) de la Península Ibérica con el propósito de detectar una posible diferenciación genética entre ambas que puede relacionarse con la diferente latitud de dichas poblaciones.

3- Estudiar el nivel y el patrón del polimorfismo nucleotídico de los genes *Fst* y *Dca* de *D. subobscura* y compararlo con el de *D. melanogaster* con el objetivo de detectar posibles semejanzas que muestren una posible acción común de la selección natural en ambas especies.

4- Estudiar el nivel y el patrón del polimorfismo nucleotídico en poblaciones naturales de *D. melanogaster* y *D. subobscura* del gen *RC* que, aunque no se ha relacionado directamente con la resistencia al frío, es un gen parálogo del gen *Dca*. El objetivo de este estudio es comparar la variabilidad intraespecífica entre los genes *Dca* y *RC* y detectar la posible acción de la selección natural en la evolución del gen *RC*.

5- Estudiar el nivel y el patrón del polimorfismo nucleotídico en poblaciones naturales de *D. melanogaster* del gen *CG6296*. Este gen comparte con el gen *Fst* la presencia de unas repeticiones internas en la mitad terminal de la región codificadora. El

objetivo principal de este estudio es analizar la variabilidad intraespecífica surgida por adiciones/deleciones en el número de estas repeticiones internas.

6- Estudiar la divergencia de los genes *Fst* y *Dca* en el subgénero *Sophophora* de *Drosophila* con el objetivo de detectar posibles aceleraciones en la tasa de sustitución no sinónima indicativas de cambios adaptativos en determinados linajes.

7- Estudiar la divergencia de los genes parálogos *Dca* y *RC* en el subgénero *Sophophora* de *Drosophila* tras su caracterización en diversas especies de los grupos *melanogaster* y *obscura*. El principal objetivo de este estudio es intentar detectar como la posible adquisición de una nueva función ha afectado a la evolución molecular de estos genes.

8- Estudiar la divergencia de los genes *Fst* y *CG6296* en el subgénero *Sophophora* de *Drosophila* tras su caracterización en diversas especies de los grupos *melanogaster* y *obscura*. El principal objetivo de este estudio es analizar la evolución molecular de las repeticiones internas presentes en estos genes.

MATERIALES Y MÉTODOS

1. MUESTRA POBLACIONAL

Las poblaciones naturales de *D. melanogaster* analizadas se capturaron en Montemayor (M) y Sant Sadurní d'Anoia (CN), localidades españolas ubicadas en Córdoba y Barcelona, respectivamente. A partir de los individuos del campo se establecieron líneas isocromosómicas para el cromosoma 3 mediante cruzamientos adecuados utilizando la cepa de letales equilibrados TM6/MKRS para la población de Montemayor (Cirera y Aguadé 1997; Ramos-Onsins y Aguadé 1998) y la TM6B/MKRS para la población de Sant Sadurní d'Anoia (Orengo y Aguadé, comunicación personal). Para esta última población también se obtuvieron líneas isocromosómicas del cromosoma X empleando la cepa de letales equilibrados FM6, $y^{31d} sc^8 dm B/1(1)X10$ (Orengo y Aguadé 2004). Las poblaciones naturales analizadas de *D. subobscura* fueron capturadas en Galicia (El Pedroso) y Cataluña (Riba-roja d'Ebre). En el caso de la población gallega (Rozas et al. 1995), se establecieron líneas isocromosómicas para el cromosoma O utilizando la cepa de letales equilibrados *Va/Ba*. En el caso de la población catalana (Munté et al. 2000) se analizaron machos individuales recién capturados o descendientes de líneas isohembras. Estos machos se habían cruzado previamente con hembras de la cepa *chcu* y congelado a -80 °C una vez asegurada su descendencia. El análisis citogenético de una larva hembra de dicha descendencia permitió determinar la ordenación cromosómica para el cromosoma X del macho analizado.

Las líneas de *D. mauritiana*, *D. yakuba* y *D. erecta* fueron gentilmente cedidas por F. Lemeunier. A partir de ellas se establecieron líneas altamente consanguíneas tras 10 generaciones sucesivas de apareamientos hermano-hermana. Se disponía en el laboratorio de líneas altamente consanguíneas obtenidas por la misma metodología de las especies *D. madeirensis* (endémica de la isla de Madeira), *D. guanche* (endémica de las Islas Canarias) y *D. simulans*. Esta última se obtuvo a partir de una línea isohembra de una población natural de Montblanc (Tarragona) (Aguadé, comunicación personal).

2. EXTRACCIÓN DE ADN

La extracción de ADN genómico de las líneas consanguíneas de las especies *D. simulans*, *D. mauritiana*, *D. yakuba*, *D. erecta*, *D. madeirensis* y *D. guanche* se realizó usando una modificación del protocolo 48 de Ashburner (1989a). Para las líneas de *D.*

melanogaster de las poblaciones de Montemayor y Sant Sadurní d'Anoia así como para las de *D. subobscura* de la población de Galicia ya se disponía de ADN genómico purificado. La extracción de ADN genómico de los machos individuales de *D. subobscura* de la población de Riba-roja d'Ebre se efectuó con el *Kit Puregene DNA purification system* (Gentra systems).

3. AMPLIFICACIÓN Y SECUENCIACIÓN

La secuencia nucleotídica de la región genómica correspondiente a los genes *Fst*, *Dca*, *Regucalcin* y CG6296 se obtuvo a partir del banco de datos del proyecto de secuenciación del genoma de *Drosophila melanogaster* con los códigos de acceso AE003683, AE003706, AE003487 y AE003758, respectivamente. Los datos accesibles del proyecto de secuenciación del genoma de *D. pseudoobscura* permitieron identificar en esta especie las regiones genómicas que contienen los genes *Fst*, *Dca* y *Regucalcin* tras realizar búsquedas BLAST (<http://www.hgsc.bcm.tmc.edu/blast/?organism=Dpseudoobscura>) a partir de las secuencias nucleotídicas de *D. melanogaster*. Una región de aproximadamente 3000 pb que contenía más o menos centrada la región codificadora de cada uno de los genes fue escogida con el fin de diseñar *primers* de amplificación por PCR. Para cada gen se diseñaron dos juegos de *primers* de amplificación de 21 pb. Los *primers* diseñados sobre las secuencias de *D. melanogaster* y *D. pseudoobscura* fueron empleados para amplificar los diferentes genes en *D. melanogaster* y *D. subobscura*, respectivamente.

Las pruebas de amplificación por PCR se realizaron en reacciones de 25 µl de volumen final. Cada reacción contenía: 18.375 µl de H₂O, 2.5 µl de Buffer 10x (*Pharmacia*), 2.0 µl de dNTP'S (2.5 nM de cada desoxinucleótido), 0.5 µl de primer *forward* (100 ng/µl), 0.5 µl de primer *reverse* (100 ng/µl), 0.125 µl Taq polimerasa (*Pharmacia*, 5 unidades /µl) y 1.0 µl de ADN genómico. Se realizaron varias pruebas con las cuatro combinaciones posibles de *primers* a temperaturas de asociación que variaron entre los 47°C y 55°C. Las diferentes pruebas permitieron establecer el juego de *primers* más idóneo para cada especie (tabla 1).

Tabla 1. *Primers de amplificación y secuenciación de los genes Fst, Dca, RC y CG6296 para cada especie.*

Gen	Especie	Amplificación		Secuenciación	
		Forward / Reverse	Forward	Reverse	Reverse
<i>Fst</i>	<i>D. melanogaster</i>	FstA, 5' GCATTGTCTCCGTTGTCTG 3'	FstA, 5' GCATTGTCTCCGTTGTCTG 3'	FstB, 5' TGTCACGCCTCCGCAATTTA 3'	
		FstB, 5' TGTCACGCCTCCGCAATTTA 3'	Fst2, 5' GTCTGGACAGGTGCTCTTTG 3'	FstR1, 5' GATGAATCCACAAGTGAG 3'	
	<i>D. simulans</i>	FstF2, 5' GTCTGGACAGGTGCTCTTTG 3'	FstF2, 5' GTCTGGACAGGTGCTCTTTG 3'	FstR0, 5' AGAAGAGCGTAAAACCAIT 3'	
		FstR0, 5' AGAAGAGCGTAAAACCAIT 3'	FstF3, 5' GGAACAGAGGTGGAATAAG 3'	FstR1, 5' GATGAATCCACAAGTGAG 3'	
	<i>D. mauritiana</i>	FstF1, 5' GCTTATCGGTGTGTGTT 3'	FstF1, 5' GCTTATCGGTGTGTGTT 3'	FstR2, 5' GAGTTGACCACCGAGGAT 3'	
		FstA, 5' ATTCTGCCCCATCAJAAAAAC 3'	FstF2, 5' GTCTGGACAGGTGCTCTTTG 3'	FstR3, 5' ATGCTGGCAGTGAATGG 3'	
	<i>D. yakuba</i>	FstF1, 5' GCTTATCGGTGTGTGTT 3'	FstF1, 5' GCTTATCGGTGTGTGTT 3'	FstR3, 5' ATGCTGGCAGTGAATGG 3'	
		FstB, 5' TGTCACGCCTCCGCAATTTA 3'	FstF2, 5' GTCTGGACAGGTGCTCTTTG 3'	FstA, 5' ATTCTGCCCCATCAJAAAAAC 3'	
	<i>D. erecta</i>	FstF1, 5' GCTTATCGGTGTGTGTT 3'	FstF1, 5' GCTTATCGGTGTGTGTT 3'	YakR1N, 5' CGAAGAATCCACCACCGA 3'	
		FstB, 5' TGTCACGCCTCCGCAATTTA 3'	FstF2, 5' GTCTGGACAGGTGCTCTTTG 3'	YakR2N, 5' TGACCAATCCCGCAGAGG 3'	
	<i>D. subobscura</i>	PstA, 5' ACAATGGACAGGGACGGTTG 3'	PstA, 5' ACAATGGACAGGGACGGTTG 3'	FstR3, 5' ATGCTGGCAGTGAATGG 3'	
		PstA, 5' ACAATGGACAGGGACGGTTG 3'	Fin, 5' TGCTACCGATTAGATAAG 3'	FstB, 5' TGTCACGCCTCCGCAATTTA 3'	
	<i>D. muleirensis</i>	PstA, 5' ACAATGGACAGGGACGGTTG 3'	PstA, 5' ACAATGGACAGGGACGGTTG 3'	EreR2N, 5' AACGAATCTGCTGAGGT 3'	
		PstA, 5' ACAATGGACAGGGACGGTTG 3'	MadF2N, 5' CCACCACCAATCTCTG 3'	FstR3, 5' ATGCTGGCAGTGAATGG 3'	
	<i>D. guanche</i>	PstA, 5' ACAATGGACAGGGACGGTTG 3'	PstA, 5' ACAATGGACAGGGACGGTTG 3'	FstB, 5' TGTCACGCCTCCGCAATTTA 3'	
		PstA, 5' ACAATGGACAGGGACGGTTG 3'	Guaf1N, 5' CAGAGATGCTTTAACCA 3'	EreR2N, 5' AACGAATCTGCTGAGGT 3'	
<i>Dca</i>	<i>D. melanogaster</i>	Dca5D, 5' TTTGAGAAATGCCATCGTGTA 3'	Dca5D, 5' TTTGAGAAATGCCATCGTGTA 3'	Dca3B, 5' GCTTCGGCACATTTTAGATA 3'	
		Dca3B, 5' GCTTCGGCACATTTTAGATA 3'	DcaF1, 5' GCTGACCGAGAAGTATC 3'	DcaR1, 5' CGGTGCTTGATGTAGT 3'	
	<i>D. simulans</i>	Dca5C, 5' AGTCGGGTCAGAGGCACAGTT 3'	DcaF1, 5' GCTGACCGAGAAGTATC 3'	DcaR2, 5' CGTAGGAATCGGGAACAG 3'	
		Dca3B, 5' GCTTCGGCACATTTTAGATA 3'	DcaF2, 5' AGGAGTTTCCGTAAGGAT 3'	DcaR4, 5' CGAATAAGCCAGAGAGAT 3'	
	<i>D. mauritiana</i>	Dca5C, 5' AGTCGGGTCAGAGGCACAGTT 3'	DcaF1, 5' GCTGACCGAGAAGTATC 3'	Dca3B, 5' GCTTCGGCACATTTTAGATA 3'	
		Dca3A, 5' AATCTGAATCGTTTGCCCTAC 3'	DcaF2, 5' AGGAGTTTCCGTAAGGAT 3'	DcaR1, 5' CGGTGCTTGATGTAGT 3'	
	<i>D. yakuba</i>	Dca5C, 5' AGTCGGGTCAGAGGCACAGTT 3'	DcaF1, 5' GCTGACCGAGAAGTATC 3'	DcaR2, 5' CGTAGGAATCGGGAACAG 3'	
		Dca3A, 5' AATCTGAATCGTTTGCCCTAC 3'	DcaF2, 5' AGGAGTTTCCGTAAGGAT 3'	DcaR4, 5' CGAATAAGCCAGAGAGAT 3'	
	<i>D. erecta</i>	Dca5C, 5' AGTCGGGTCAGAGGCACAGTT 3'	DcaF1, 5' GCTGACCGAGAAGTATC 3'	Dca3A, 5' AATCTGAATCGTTTGCCCTAC 3'	
		Dca3A, 5' AATCTGAATCGTTTGCCCTAC 3'	DcaF2, 5' AGGAGTTTCCGTAAGGAT 3'	Dca3B, 5' GCTTCGGCACATTTTAGATA 3'	
	<i>D. subobscura</i>	PstA, 5' CCGTCATCACCGTTAAGGTT 3'	DcaF3, 5' CGTCTTCAAGGTCAATCC 3'	DcaR1, 5' CGGTGCTTGATGTAGT 3'	
		PstA, 5' CCGTCATCACCGTTAAGGTT 3'	DcaF4, 5' GGTATGACCGTAGATACC 3'	DcaR2, 5' CGTAGGAATCGGGAACAG 3'	
	<i>D. muleirensis</i>	PstA, 5' CCGTCATCACCGTTAAGGTT 3'	Dca5C, 5' AGTCGGGTCAGAGGCACAGTT 3'	DcaR4, 5' CGAATAAGCCAGAGAGAT 3'	
		PstA, 5' CCGTCATCACCGTTAAGGTT 3'	EreF1N, 5' AGCAATTCGGGACATGA 3'	Dca3A, 5' AATCTGAATCGTTTGCCCTAC 3'	
	<i>D. guanche</i>	PstA, 5' CCGTCATCACCGTTAAGGTT 3'	EreF1N, 5' AGCAATTCGGGACATGA 3'	Dca3B, 5' GCTTCGGCACATTTTAGATA 3'	
		PstA, 5' CCGTCATCACCGTTAAGGTT 3'	DcaF2, 5' AGGAGTTTCCGTAAGGAT 3'	DcaR1, 5' CGGTGCTTGATGTAGT 3'	
<i>D. melanogaster</i>	PstA, 5' CCGTCATCACCGTTAAGGTT 3'	EreF2N, 5' AGCTGACCGAGAAGTATC 3'	DcaR2, 5' CGTAGGAATCGGGAACAG 3'		
	PstA, 5' CCGTCATCACCGTTAAGGTT 3'	DcaF4, 5' CGTCTTCAAGGTCAATCC 3'	DcaR4, 5' CGAATAAGCCAGAGAGAT 3'		
<i>D. simulans</i>	PstA, 5' CCGTCATCACCGTTAAGGTT 3'	EreR1N, 5' ACAGCACTCTCGATCAA 3'	Dca3A, 5' AATCTGAATCGTTTGCCCTAC 3'		
	PstA, 5' CCGTCATCACCGTTAAGGTT 3'	EreR2N, 5' GCACAACTGGAACACT 3'	Dca3B, 5' GCTTCGGCACATTTTAGATA 3'		
<i>D. mauritiana</i>	PstA, 5' CCGTCATCACCGTTAAGGTT 3'	EreR3N, 5' ATCGGTTTGGCATACT 3'	DcaR1, 5' CGGTGCTTGATGTAGT 3'		
	PstA, 5' CCGTCATCACCGTTAAGGTT 3'	EreR4N, 5' GGTAATTTGTTGGTGG 3'	DcaR2, 5' CGTAGGAATCGGGAACAG 3'		
<i>D. yakuba</i>	PstA, 5' CCGTCATCACCGTTAAGGTT 3'	EreR5N, 5' AACCAAGTGGCTTGATCA 3'	DcaR4, 5' CGAATAAGCCAGAGAGAT 3'		
	PstA, 5' CCGTCATCACCGTTAAGGTT 3'	PstA, 5' CCGTCATCACCGTTAAGGTT 3'	Dca3A, 5' AATCTGAATCGTTTGCCCTAC 3'		
<i>D. erecta</i>	PstA, 5' CCGTCATCACCGTTAAGGTT 3'	Sngf1N, 5' CGCTGCCGACTCTTAIG 3'	Dca3B, 5' GCTTCGGCACATTTTAGATA 3'		
	PstA, 5' CCGTCATCACCGTTAAGGTT 3'	Sngf2N, 5' CCAATCCAAGGTCGTC 3'	DcaR1, 5' CGGTGCTTGATGTAGT 3'		
<i>D. subobscura</i>	PstA, 5' CCGTCATCACCGTTAAGGTT 3'	PstA, 5' CCGTCATCACCGTTAAGGTT 3'	DcaR2, 5' CGTAGGAATCGGGAACAG 3'		
	PstA, 5' CCGTCATCACCGTTAAGGTT 3'	Sngf1N, 5' CGCTGCCGACTCTTAIG 3'	DcaR4, 5' CGAATAAGCCAGAGAGAT 3'		
<i>D. muleirensis</i>	PstA, 5' CCGTCATCACCGTTAAGGTT 3'	Sngf2N, 5' CCAATCCAAGGTCGTC 3'	Dca3A, 5' AATCTGAATCGTTTGCCCTAC 3'		
	PstA, 5' CCGTCATCACCGTTAAGGTT 3'	PstA, 5' CCGTCATCACCGTTAAGGTT 3'	Dca3B, 5' GCTTCGGCACATTTTAGATA 3'		
<i>D. guanche</i>	PstA, 5' CCGTCATCACCGTTAAGGTT 3'	Sngf1N, 5' CGCTGCCGACTCTTAIG 3'	DcaR1, 5' CGGTGCTTGATGTAGT 3'		
	PstA, 5' CCGTCATCACCGTTAAGGTT 3'	Sngf2N, 5' CCAATCCAAGGTCGTC 3'	DcaR2, 5' CGTAGGAATCGGGAACAG 3'		

Tabla1. Continuación

Gen	Especie	Amplificación		Secuenciación	
		Forward / Reverse		Forward	Reverse
RC	<i>D. melanogaster</i>	Reg5B. 5' GCCCCTGTGCCCTCTACAT 3'	Reg5B. 5' GCCCCTGTGCCCTCTACAT 3'	Reg3B. 5' ACCAAGGAGCCAACCCCGT 3'	RegR1. 5' TTTTACGCCACCTTGCTA 3'
		Reg3B. 5' ACCAAGGAGCCAACCCCGT 3'	RegF1. 5' GCCGATCTCAGCCAGTA 3'	RegR2. 5' ATTTACTCGGCTAAGG 3'	RegR3. 5' CCTCGTAGTCGGTGGTAT 3'
			RegF2. 5' GGC AACCATGATAATGAT 3'	RegR4. 5' CCGATTGCGGATGATGC 3'	RegR5. 5' CAATCACGGGAATCAGTA 3'
			RegF3. 5' CCCGTGGTCGCTTCTTG 3'		
			RegF4. 5' TTACCTCCGCGCCTTCG 3'		
	<i>D. subobscura</i>	Ps5B. 5' GCTCTCTTTGGCTTGACGGTA 3'	Ps5B. 5' GCTCTCTTTGGCTTGACGGTA 3'	Ps3B. 5' GCGATGATGACGATGATGACT 3'	SubR1N. 5' GCTACACTTTTGGCTTCA 3'
		Ps3B. 5' GCGATGATGACGATGATGACT 3'	SubF1N. 5' GGAATGACGCCAAGACTG 3'		SubR2N. 5' GCTTGGTGGAAACTTGA 3'
			SubF2N. 5' ACAGAAATGGCGAGGTCC 3'		SubR3N. 5' CAGTCTTGGCGTCAATCC 3'
	<i>D. madeirensis</i>	Ps5B. 5' GCTCTCTTTGGCTTGACGGTA 3'	Ps5B. 5' GCTCTCTTTGGCTTGACGGTA 3'	MadF1N. 5' GTAGGTGCTGGATAGCC 3'	MadR1N. 5' AACTCGCTGGCTTTGTC 3'
		Ps3B. 5' GCGATGATGACGATGATGACT 3'	MadF2N. 5' TCGTAGTCGGTGGTGGTCG 3'		MadR2N. 5' CAAAACAGCCCAAGGA 3'
	<i>D. guanche</i>	Ps5B. 5' GCTCTCTTTGGCTTGACGGTA 3'	Ps5B. 5' GCTCTCTTTGGCTTGACGGTA 3'	GuaF1N. 5' GCAGCGTGGAAATGGAT 3'	GuaR1N. 5' TGAAGCCAAAAGTGTAGC 3'
		Ps3B. 5' GCGATGATGACGATGATGACT 3'	SubF1N. 5' GGAATGACGCCAAGACTG 3'	SubF2N. 5' ACAGAAATGGCGAGGTCC 3'	GuaR2N. 5' TGCGAAAAGGAGCCGAGG 3'
				GuaR3N. 5' GTAGGTGCTGGATAGCC 3'	
CG6296	<i>D. melanogaster</i>	Cg5A. 5' CTACGGAATGGGCAACTAAGT 3'	Cg5A. 5' CTACGGAATGGGCAACTAAGT 3'	Cg3B. 5' GTGACTGGCGTTAATGTG 3'	CgR1. 5' CAGCGGTTGTGGATACAT 3'
		Cg3B. 5' GTGACTGGCGTTAATGTG 3'	CgF1. 5' TGTGCTGGGCTGCGTGAT 3'	CgR2. 5' GGTCTCTTCGGGCTTCTC 3'	CgR3. 5' ACGGAGGATCCGACTTT 3'
			CgF2. 5' ACAAGGATGGGTTGAACA 3'	CgR4. 5' CCACAGAGGAAGCGAAT 3'	CgR5. 5' GCTAGGTTACCGAAATCT 3'
			CgF3. 5' CAGCCTGGTTGTGGTATC 3'		
			CgF4. 5' ACGGAAGCCCAAGAAC 3'		
	<i>D. simulans</i>	Cg5A. 5' CTACGGAATGGGCAACTAAGT 3'	Cg5A. 5' CTACGGAATGGGCAACTAAGT 3'	Cg3A. 5' GGGAAAGTAGCCATTGGTAAC 3'	SimR2N. 5' GTTGTAGTAGTCGAGTCT 3'
		Cg3A. 5' GGGAAAGTAGCCATTGGTAAC 3'	CgF1. 5' TGTGCTGGGCTGCGTGAT 3'		SimR4N. 5' TCCCACCTTGCAGAGTT 3'
			CgF2. 5' ACAAGGATGGGTTGAACA 3'	MauR4N. 5' CCACAGAGGTAGCGTACT 3'	
			CgF3. 5' CAGCCTGGTTGTGGTATC 3'	CgR5. 5' GCTAGGTTACCGAAATCT 3'	
			CgF5. 5' CTACAACAACCGCAGAAC 3'		
	<i>D. mauritiana</i>	Cg5B. 5' TGTGAGCGAAAGCAATCC 3'	Cg5B. 5' TGTGAGCGAAAGCAATCC 3'	Cg3A. 5' GGGAAAGTAGCCATTGGTAAC 3'	MauR1N. 5' CAGGGTTGTGGATACAT 3'
		Cg3A. 5' GGGAAAGTAGCCATTGGTAAC 3'	CgF1. 5' TGTGCTGGGCTGCGTGAT 3'		MauR2N. 5' GGTCTCTTCGGGCTTTC 3'
			CgF2. 5' ACAAGGATGGGTTGAACA 3'	MauR3N. 5' GGTGGATTCGCATCTCAI 3'	
			CgF3. 5' CAGCCTGGTTGTGGTATC 3'	MauR4N. 5' CCACAGAGGTAGCGTACT 3'	
			MauF4N. 5' AAGAAGACCCCAAGGCA 3'	MauR5N. 5' GCACAAGAAGGCACAATC 3'	
			CgF5. 5' CTACAACAACCGCAGAAC 3'	MauF6N. 5' GCAACTGGCGAAGATGAA 3'	
	<i>D. yakuba</i>	Cg5B. 5' TGTGAGCGAAAGCAATCC 3'	CgF1. 5' TGTGCTGGGCTGCGTGAT 3'	Cg3A. 5' GGGAAAGTAGCCATTGGTAAC 3'	YakR1N. 5' CTCATTGGGAACTGTGG 3'
		Cg3A. 5' GGGAAAGTAGCCATTGGTAAC 3'	CgF2. 5' ACAAGGATGGGTTGAACA 3'		YakR2N. 5' TGTCTGAAGGAAGAGTG 3'
			CgF3. 5' CAGCCTGGTTGTGGTATC 3'	MauF4N. 5' AAGAAGACCCCAAGGCA 3'	YakR3N. 5' ACCCTCTGCGACAAGAA 3'
			MauF5N. 5' CTACAACAACCGCAGAAC 3'	YakF5N. 5' CTACAACAACCGCAGAAC 3'	YakR4N. 5' GCTAGGTTACCGAAACT 3'
	<i>D. erecta</i>	Cg5A. 5' CTACGGAATGGGCAACTAAGT 3'	Cg5A. 5' CTACGGAATGGGCAACTAAGT 3'	Cg3A. 5' GGGAAAGTAGCCATTGGTAAC 3'	EreR1N. 5' GGTGTTTCCATCTTTCGG 3'
		Cg3A. 5' GGGAAAGTAGCCATTGGTAAC 3'	CgF1. 5' TGTGCTGGGCTGCGTGAT 3'		EreR2N. 5' CCTCCTCTGGCTTATCTG 3'
			CgF2. 5' ACAAGGATGGGTTGAACA 3'	EreR3N. 5' GCATACCCATCAATACT 3'	
		CgF3. 5' CAGCCTGGTTGTGGTATC 3'	EreR5N. 5' ACCGAAAACATCAATGT 3'		
		EreF4N. 5' ACGGAAGCCTCAGAAGTC 3'	EreF5N. 5' GGACACGACTACTGAAGAGC 3'		

N, primers diseñados sobre la secuencia parcial de la respectiva especie.

Las diferentes pruebas también permitieron determinar las condiciones óptimas para la amplificación de los genes en las líneas de *D. melanogaster* y *D. subobscura* (tablas 2 y 3).

Tabla 2. Condiciones de amplificación por PCR para las especies del subgrupo *melanogaster*.

Gen	PCR	<i>D. melanogaster</i>	<i>D. simulans</i>	<i>D. mauritiana</i>	<i>D. yakuba</i>	<i>D. erecta</i>	
<i>Fst</i>	Desnaturalización	94°C 3min	94°C 3min	94°C 3min	94°C 3min	94°C 3min	
	Asociación	55°C 30seg	48°C 30seg	50°C 30seg	48°C 30seg	48°C 30seg	
	Extensión	68°C 2.5min	68°C 2min	68°C 2.5min	68°C 2.5min	68°C 2.5min	
	Desnaturalización	92°C 30seg	92°C 30seg	92°C 30seg	92°C 30seg	92°C 30seg	
	Asociación	55°C 30seg	48°C 30seg	50°C 30seg	48°C 30seg	48°C 30seg	
	Extensión	68°C 2.5min	68°C 2min	68°C 2.5min	68°C 2.5min	68°C 2.5min	
	Desnaturalización	92°C 30seg	92°C 30seg	92°C 30seg	92°C 30seg	92°C 30seg	
	Asociación	55°C 30seg	48°C 30seg	50°C 30seg	48°C 30seg	48°C 30seg	
	Extensión	68°C 5min	68°C 4min	68°C 5min	68°C 5min	68°C 5min	
	<i>Dca</i>	Desnaturalización	94°C 3min	94°C 3min	94°C 3min	94°C 3min	94°C 3min
		Asociación	52°C 30seg	54°C 30seg	52°C 30seg	54°C 30seg	54°C 30seg
		Extensión	68°C 3min	68°C 3min	68°C 3min	68°C 3min	68°C 3min
Desnaturalización		92°C 30seg	92°C 30seg	92°C 30seg	92°C 30seg	92°C 30seg	
Asociación		52°C 30seg	54°C 30seg	52°C 30seg	54°C 30seg	54°C 30seg	
Extensión		68°C 3min	68°C 3min	68°C 3min	68°C 3min	68°C 3min	
Desnaturalización		92°C 30seg	92°C 30seg	92°C 30seg	92°C 30seg	92°C 30seg	
Asociación		52°C 30seg	54°C 30seg	52°C 30seg	54°C 30seg	54°C 30seg	
Extensión		68°C 6min	68°C 6min	68°C 6min	68°C 6min	68°C 6min	
<i>CG6296</i>		Desnaturalización	94°C 3min	94°C 3min	94°C 3min	94°C 3min	94°C 3min
		Asociación	50°C 30seg	48°C 30seg	48°C 30seg	48°C 30seg	48°C 30seg
		Extensión	68°C 3.5min	68°C 3.5min	68°C 3.5min	68°C 3.5min	68°C 3.5min
	Desnaturalización	92°C 30seg	92°C 30seg	92°C 30seg	92°C 30seg	92°C 30seg	
	Asociación	50°C 30seg	48°C 30seg	48°C 30seg	48°C 30seg	48°C 30seg	
	Extensión	68°C 3.5min	68°C 3.5min	68°C 3.5min	68°C 3.5min	68°C 3.5min	
	Desnaturalización	92°C 30seg	92°C 30seg	92°C 30seg	92°C 30seg	92°C 30seg	
	Asociación	50°C 30seg	48°C 30seg	48°C 30seg	48°C 30seg	48°C 30seg	
	Extensión	68°C 7min	68°C 7min	68°C 7min	68°C 7min	68°C 7min	
	<i>Regucalcin</i>	Desnaturalización	94°C 3min				
		Asociación	58°C 30seg				
		Extensión	68°C 3min				
Desnaturalización		92°C 30seg					
Asociación		58°C 30seg					
Extensión		68°C 3min					
Desnaturalización		92°C 30seg					
Asociación		58°C 30seg					
Extensión		68°C 6min					

Tabla 3. Condiciones de amplificación por PCR para las especies del grupo *obscura*

Gen	PCR	<i>D. subobscura</i>	<i>D. madeirensis</i>	<i>D. guanche</i>
<i>Fst</i>	Desnaturalización	94°C 3min	94°C 3min	94°C 3min
	Asociación	1 ciclo	48°C 30seg	48°C 30seg
	Extensión		68°C 3min	68°C 3min
	Desnaturalización		92°C 30seg	92°C 30seg
	Asociación	28 ciclos	48°C 30seg	48°C 30seg
	Extensión		68°C 3min	68°C 3min
	Desnaturalización		92°C 3seg	92°C 3seg
	Asociación	1 ciclo	48°C 30seg	48°C 30seg
	Extensión		68°C 6min	68°C 6min
Desnaturalización	94°C 3min		94°C 3min	94°C 3min
<i>Dca</i>	Asociación	1 ciclo	48°C 30seg	45°C 30seg
	Extensión		68°C 3min	68°C 3min
	Desnaturalización		92°C 30seg	92°C 30seg
	Asociación	28 ciclos	48°C 30seg	45°C 30seg
	Extensión		68°C 3min	68°C 3min
	Desnaturalización		92°C 30seg	92°C 3seg
	Asociación	1 ciclo	52°C 30seg	45°C 30seg
	Extensión		68°C 6min	68°C 6min
	Desnaturalización		94°C 3min	94°C 3min
<i>Regucalcin</i>	Asociación	1 ciclo	52°C 30seg	52°C 30seg
	Extensión		68°C 2min	68°C 2min
	Desnaturalización		92°C 30seg	92°C 30seg
	Asociación	28 ciclos	52°C 30seg	52°C 30seg
	Extensión		68°C 2min	68°C 2min
	Desnaturalización		92°C 30seg	92°C 30seg
	Asociación	1 ciclo	52°C 30seg	52°C 30seg
	Extensión		68°C 4min	68°C 4min
	Desnaturalización		94°C 3min	94°C 3min

Los productos resultantes de la amplificación por PCR fueron purificados por medio de columnas *Amicon Microcon-PCR kit (Millipore)* y diluidos en 50 µl de H₂O. Ambas cadenas de los fragmentos amplificados se secuenciaron por medio de *primers* internos de secuenciación diseñados aproximadamente cada 400 pb (tabla 1). Las reacciones de secuenciación se realizaron con el *kit* comercial *BigDyes3.0 (Applied*

Biosystems) siguiendo las instrucciones de la compañía y según las siguientes condiciones:

94°C 3min		
96°C 10seg		
52°C 5seg		25 ciclos
60°C 4min		
4°C ∞		

Los productos de secuenciación se recuperaron siguiendo el protocolo de precipitación con etanol de acuerdo a las recomendaciones de la casa comercial (*Applied Biosystems*) y se secaron al vacío. Los productos resultantes se analizaron con el secuenciador automático ABI 377 o ABI 3700 (*Applied Biosystems*) de los *Serveis Científico-Tècnics* de la Universitat de Barcelona.

Asimismo, se realizaron distintas pruebas con diferentes combinaciones de *primers* y condiciones para amplificar las regiones de interés en las especies del *subgrupo melanogaster* y del *grupo obscura*. Las reacciones de amplificación se efectuaron del mismo modo que las descritas para *D. melanogaster* y *D. subobscura* pero variando las condiciones de PCR para cada especie (tabla 2). Además, fue necesario diseñar *primers* específicos de secuenciación para algunas de las especies a partir de las secuencias parciales obtenidas de cada una de ellas. Las condiciones de las reacciones de secuenciación fueron las ya especificadas anteriormente.

Las secuencias nucleotídicas de las diferentes regiones en las especies *D. pseudoobscura*, *D. sechellia*, *D. ananassae*, *D. persimilis*, *D. virilis*, *D. mojavensis* y *D. grimshawi* fueron obtenidas a partir de la base de datos del proyecto de secuenciación de cada especie (<http://insects.eugenes.org/species/blast>).

4. ENSAMBLAJE Y ALINEAMIENTO

La secuencia consenso de cada una de las diferentes líneas y especies se obtuvo por medio del programa *SeqMan* (DNASar, *Lasergene*). Posteriormente, las diferentes secuencias consenso de las líneas de *D. melanogaster* y *D. subobscura* se alinearon con

el programa *Clustal W* versión 1.8 (Thompson et al. 1997). Para el alineamiento de las secuencias consenso de las diferentes especies se utilizó el programa *ProbCons* versión 1.08 (Do et al. 2005). En el caso del alineamiento interespecífico de los genes *Fst* y *CG6296* fue necesario emplear un *script* que permitiera obtener el alineamiento nucleotídico más correcto a partir del conseguido a nivel de aminoácidos. La edición de las secuencias para los análisis posteriores fue realizada con el programa *MacClade* versión 4.05 (Maddison y Maddison 1989). Los análisis de *DotPlot* para identificar las regiones de alta similitud de los genes *Dca* y *Regucalcin* se realizaron con la aplicación implementada en el programa *MegAlign* (DNAStar, *Lasergene*).

5. HIBRIDACIÓN *IN SITU*

La localización citológica de los genes *Fst*, *Dca* y *Regucalcin* en *D. subobscura* fue determinada mediante hibridación *in situ* de acuerdo al protocolo de Segarra y Aguadé (1992) y de Montgomery et al (1987). Las sondas utilizadas fueron obtenidas tras amplificar por PCR la región codificadora de cada uno de los genes en *D. pseudoobscura*.

6. ANÁLISIS INTRAESPECÍFICOS E INTERESPECÍFICOS

El programa DnaSP version 4.20 (Rozas et al. 2003) fue utilizado para realizar el análisis de la variabilidad intraespecífica y para llevar a cabo los diferentes tests de neutralidad. El programa Mega2 (Kumar et al. 2004) se empleó para estimar la divergencia interespecífica. Esta estima se obtuvo independientemente para las posiciones sinónimas (K_s) y para las no sinónimas (K_a) de acuerdo al método de Nei y Gojobori (1986) modificado por Nei y Kumar (2000). Este mismo programa fue utilizado para obtener las genealogías de los alelos y realizar las reconstrucciones filogenéticas con el método de *neighbor-joining* (Saitou y Nei 1987).

El sesgo en el uso de codones sinónimos (*codon bias*) presente en los genes estudiados en las diferentes especies se estimó a partir de tres parámetros diferentes: χ^2 ponderada, CBI y ENC. Valores altos de la χ^2 ponderada (Shields et al. 1988) y del índice de *codon bias* (CBI) (Morton 1993) indican una fuerte desviación respecto al uso aleatorio de codones sinónimos. Por el contrario, el número efectivo de codones (ENC)

oscila entre 61 cuando todos los codones sinónimos son equifrecuentes y 20 cuando se usa únicamente un codón en cada clase de codones sinónimos (Wright 1990).

El *relative-rate test* (Sarich y Wilson 1973) se utilizó para identificar posibles desviaciones de la hipótesis del reloj molecular debido a aceleraciones en la tasa de sustitución nucleotídica en linajes particulares. Este test compara las tasas de sustitución en dos linajes y requiere información de una tercera especie de referencia o especie *outgroup*. Los métodos propuestos por Wu y Li (1985) y Tajima (1993) se utilizaron para realizar el *relative-rate test*.

7. ESTIMAS DE LA VARIABILIDAD NUCLEOTÍDICA INTRAESPECÍFICA

La variabilidad nucleotídica intraespecífica de la muestra de secuencias de ADN se estimó a partir del número de posiciones segregantes o polimórficas (S) y del número promedio de diferencias nucleotídicas (k) entre las secuencias analizadas. Este último parámetro (k) depende del número de posiciones polimórficas y de la frecuencia de las variantes nucleotídicas que presentan, pero no depende del tamaño muestral. El número promedio de diferencias nucleotídicas se define como:

$$k(\text{por secuencia}) = \frac{\sum_{ij} k_{ij}}{\binom{n}{2}}$$

donde k_{ij} es el número de diferencias nucleotídicas entre la secuencia i y la secuencia j ($i < j$), n es el número de secuencias de ADN analizadas y $\binom{n}{2} = \frac{n(n-1)}{2}$ es el número total de comparaciones entre secuencias. Si se corrige por el número de posiciones de la región analizada (L), se obtiene k (por posición), es decir el número promedio de diferencias entre secuencias por posición.

$$k(\text{por posición}) = \frac{k}{L}$$

Esta estima corresponde a la diversidad nucleotídica (π) propuesta por Nei (1987).

$$\Pi = \left(\frac{n}{n-1} \right) \sum_{ij} x_i x_j k_{ij}$$

donde n es el número de secuencias de ADN analizadas, x_i y x_j son las frecuencias de las secuencias i y j en la muestra, y k_{ij} es la proporción de nucleótidos diferentes entre las secuencias i y j .

Otro parámetro que se utiliza para describir la variabilidad nucleotídica es θ . Este parámetro se estima a partir del número de posiciones segregantes (S) en una muestra de n secuencias de ADN. Este estimador corresponde a k (por secuencia) cuando las mutaciones son selectivamente neutras, la población es panmítica y se encuentra en equilibrio mutación-deriva.

$$\theta (\text{por secuencia}) = \frac{S}{\sum_{i=1}^{n-1} \frac{1}{i}}$$

donde $\theta = 4N_e\mu$, siendo N_e el tamaño efectivo de la población y μ la tasa de mutación neutra por secuencia y por generación.

Cuando se considera el número de posiciones de la región analizada (L) se obtiene la estima de la heterozigosis por nucleótido (θ o estimador de Watterson) (Watterson 1975).

$$\theta (\text{por posición}) = \frac{s/L}{\sum_{i=1}^{n-1} \frac{1}{i}}$$

8. DIFERENCIACIÓN GENÉTICA

La diferenciación genética entre las dos poblaciones analizadas de *D. melanogaster* se estimó utilizando el estadístico F_{st} (Hudson et al. 1992) que se basa en el número promedio de diferencias entre secuencias de la misma población y el número promedio de diferencias entre secuencias de dos poblaciones diferentes. Para obtener la probabilidad asociada al valor de F_{st} observado bajo la hipótesis nula de no diferenciación ($H_0: F_{st} = 0$) se realizaron simulaciones de coalescencia asumiendo el modelo neutro en una población panmítica sin subdivisión. Se utilizaron valores de θ

compatibles con el número observado de sitios segregantes (*rejection algorithm*; Tavaré et al. 1997). Para cada una de las 10000 iteraciones realizadas se escogieron aleatoriamente dos grupos de muestras y se calcularon sus valores de F_{st} . La probabilidad del valor F_{st} observado se obtuvo a partir de la distribución de los valores de F_{st} obtenidos por simulación.

9. ESTIMAS DEL DESEQUILIBRIO DE LIGAMIENTO

El grado de la asociación entre las variantes nucleotídicas de las posiciones polimórficas se estimó mediante el parámetro D' (Lewontin 1964) que corresponde al cociente entre el valor de desequilibrio (D) y el valor de desequilibrio máximo (D_{max}). La aplicación de un test de χ^2 (Zapata y Álvarez 1993) permite determinar si el desequilibrio detectado es significativamente diferente de 0. Además, el grado de desequilibrio de ligamiento global de una región fue determinado por los estadísticos Z_{ns} (Kelly 1997), Wall's B y Q (Wall 1999) y Z_A (Rozas et al. 2001). Los intervalos de confianza de estos estadísticos se obtuvieron por simulaciones (10000 réplicas) de coalescencia asumiendo una población con un tamaño constante (Kingman 1982; Hudson 1983; Hudson 1990; Rozas y Rozas 1999). Las simulaciones de coalescencia se efectuaron condicionando al número de sitios segregantes y en general asumiendo no recombinación. En algunos casos se fijaron diferentes valores del parámetro de recombinación.

10. ESTIMAS DE LA TASA DE RECOMBINACIÓN

La tasa de recombinación promedio entre dos nucleótidos adyacentes de una secuencia de ADN, r , puede estimarse a partir del parámetro R , siendo $R = 4N_e r$ ($R = 3N_e r$, para regiones situadas en el cromosoma X) (Hudson 1987). Este estimador se basa entre otros parámetros en la varianza del número de diferencias nucleotídicas entre secuencias en comparaciones dos a dos.

11. TEST DEL NEUTRALISMO

Se aplicaron diversos test de neutralismo. El test de Tajima (1989) se basa en el estadístico D que corresponde a la diferencia normalizada entre el número promedio

observado de diferencias nucleotídicas entre secuencias (k por secuencia) y la estima de este parámetro esperada a partir del número de posiciones segregantes para mutaciones neutras en una población en equilibrio mutación-deriva (θ por secuencia). Ambas estimas deben ser iguales de acuerdo al neutralismo. Por lo tanto, asumiendo equilibrio mutación-deriva se espera que $D = 0$ si las mutaciones son neutras. Debido a que el número promedio de diferencias nucleotídicas tiene en cuenta la frecuencia observada a la que segregan las variantes nucleotídicas y el estimador de Watterson no, el estadístico D indica la posible desviación del espectro de frecuencia observado respecto al esperado para mutaciones neutras en una población en equilibrio mutación-deriva.

El test de Fu y Li (1993) se basa en las predicciones sobre el número de mutaciones esperadas en las ramas externas e internas de una genealogía neutra asumiendo equilibrio mutación-deriva. Este test puede aplicarse con *outgroup* (estadísticos D y F) o sin *outgroup* (estadísticos D^* y F^*). Los tests con *outgroup* se basan en el número de mutaciones en las ramas externas de la genealogía, mientras que los tests sin *outgroup* se basan en el número de mutaciones únicas en la muestra. Para estimar los estadísticos D^* y D se utiliza el número total de mutaciones y bien el número de mutaciones únicas en la muestra o el número de mutaciones en las ramas externas, respectivamente. Los estadísticos F^* y F se estiman a partir del número promedio de diferencias entre pares de secuencias y bien el número de mutaciones únicas en la muestra o el número de mutaciones en las ramas externas, respectivamente.

El test de Hudson, Kreitman y Aguadé (1987) se basa en la predicción del neutralismo de que la relación polimorfismo/divergencia se mantiene en diferentes regiones que evolucionan de acuerdo al neutralismo. Por lo tanto, este test compara los niveles de polimorfismo y divergencia de la región de interés con los de una región neutra de referencia.

El test de McDonald y Kreitman (1991) se basa en la comparación de la relación entre el número de diferencias no sinónimas y sinónimas fijadas entre especies con la relación entre el número de polimorfismos no sinónimos y sinónimos dentro de especie. Ambas relaciones se espera que sean similares para mutaciones estrictamente neutras. Este test no requiere que las poblaciones se encuentren en equilibrio mutación-deriva y puede aplicarse si se dispone de datos intraespecíficos en una especie y de una secuencia única en una especie cercana.

El test de Fay y Wu (2000) se basa en el estadístico H que permite detectar desviaciones en el espectro de frecuencias observado para las mutaciones derivadas

(mutaciones no ancestrales) respecto del esperado bajo el modelo neutro. La aplicación de este test requiere el conocimiento de un alelo ancestral que habitualmente se infiere a partir de la secuencia de una especie de referencia (*outgroup*).

12. ANÁLISIS DE *MAXIMUM LIKELIHOOD* (ML)

Para obtener las estimas de *ML* bajo diferentes modelos de sustitución de codones se utilizaron los modelos de *random-effects* disponibles en el programa *codeml* implementado en el paquete PAML version 3.14 (Yang 1997).

Los *branch models* M0 (*one-ratio*), FR (*Free-ratio*) y *branch-specific* permiten asumir variación en el ratio de ω ($\omega = d_N/d_S$, donde d_N y d_S son el número de sustituciones no sinónimas y sinónimas, respectivamente) entre las ramas de la filogenia y son útiles para detectar selección positiva actuando en linajes particulares (Yang 1998; Yang y Nielsen 2000). Por otro lado, los *site models* M1 (*nearly neutral*), M2 (*positive selection*), M3 (*discrete*), M7 (beta) y M8 (Beta& ω) permiten que el ratio de ω varíe entre las posiciones (entre codones o aminoácidos en la proteína) (Nielsen y Yang 1998; Yang et al. 2000). El modelo de *branch-site A* permite que el ratio de ω pueda variar tanto entre posiciones como entre diferentes linajes. Este modelo intenta detectar la acción de la selección positiva afectando únicamente a algunos aminoácidos en pocos linajes (Yang y Nielsen 2002).

El modelo M0 se aplicó para estimar la longitud de las ramas de la filogenia puesto que permite obtener los valores iniciales adecuados que son utilizados en el momento de contrastar modelos de sustitución de codones más complejos. El programa *codeml* se ejecutó varias veces para el mismo modelo con diferentes valores iniciales de ω con el objetivo de evitar una baja precisión en la identificación bayesiana de codones bajo selección positiva. El *likelihood ratio test* (LRT) fue usado para contrastar el ajuste de los datos entre dos modelos que se encuentren anidados. Este test asume que dos veces la diferencia del *log likelihood* ($2\Delta c$) de los dos modelos sigue una distribución de χ^2 con un número de grados de libertad igual a la diferencia en el número de parámetros libres de ambos modelos (Whelan y Goldman 1999). Los valores de *P* se obtuvieron utilizando el programa *chi2* también incluido en el paquete PAML.

RESULTADOS Y DISCUSIÓN

CAPÍTULO I
EL GEN FROST (*Fst*)

I.1. POLIMORFISMO NUCLEOTÍDICO DEL GEN *Fst*

I.1.1 RESULTADOS

I.1.1.1 POLIMORFISMO NUCLEOTÍDICO DEL GEN *Fst* EN *D. MELANOGASTER*

Se secuenció un fragmento de aproximadamente 1.5 kb en 30 líneas de *D. melanogaster*. Dicho fragmento incluye la región codificadora del gen *Frost* (*Fst*) que comprende alrededor de 0.9 kb y parte de las regiones flanqueantes 5' y 3' con una longitud aproximada de 0.4 kb y 0.2 kb, respectivamente.

Diferenciación genética entre poblaciones

La posible diferenciación genética para la región estudiada entre las líneas de las poblaciones de Montemayor y Sant Sadurní d'Anoia fue estimada usando el estadístico F_{st} (Hudson et al. 1992). El valor obtenido de $F_{st} = 0.0037$ no es significativamente diferente al observado bajo el modelo neutro de una población panmítica sin subdivisión (figura I.1). Dada la ausencia de diferenciación genética, ambas poblaciones se consideraron conjuntamente en los análisis posteriores.

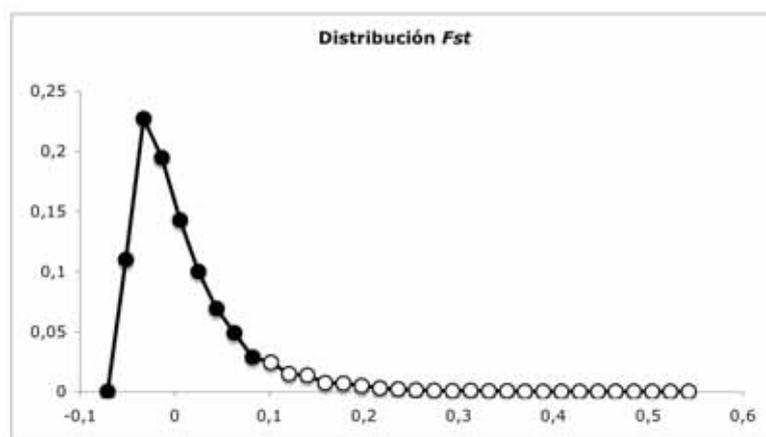


Figura I.1. Distribución del estadístico F_{st} (Hudson et al. 1992) en una población panmítica de 30 individuos (ver materiales y métodos).

Análisis del polimorfismo nucleotídico

El alineamiento múltiple de las 30 secuencias de *D. melanogaster* presenta un total de 1576 posiciones. Sin embargo, el número de posiciones alineadas se reduce a 1467 al eliminar aquellas posiciones con *gaps* en una o varias de las líneas analizadas. En el alineamiento se detectaron tanto polimorfismos por cambio nucleotídico como polimorfismos por inserción/delección. Se identificaron un total de 24 posiciones polimórficas por cambio nucleotídico y 5 polimorfismos por inserción/delección (figura I.2). La figura I.2 también incluye la secuencia previamente publicada del cDNA del gen *Fst* (Goto 2001) y la identificada en el proyecto de secuenciación del genoma de *D. melanogaster* (código de acceso AE003683).

En la región 5' flanqueante se detectaron 6 posiciones polimórficas por cambio nucleotídico de las que 4 presentan variantes únicas. En esta región también se observó un polimorfismo por inserción/delección de 4 nucleótidos. En la región 3' flanqueante, únicamente se detectaron 2 posiciones polimórficas por cambio nucleotídico, una de las cuales presenta una variante única (figura I.2).

	2	28	115	178	180	181	210	260	300	334	543	586	604	627	632	639	664	689	736	783	785	872	916	917	986	1054	1068	1125	1140	1143	1154	1175	1204	1241	1250	1311	1325	1340	1392	1399	1469	1543							
M47	A	T	G	T	G	T	G	A	G	G	T	G	A	C	A	C	A	C	A	C	A	G	T	G	G	G	T	A	A	G	C	G	-	A	C	C	G	C	A	A									
M55			
CN1	.	.	G	T	A		
CN9	
CN16	A	
CN18
CN22
CN25
CN26
CN41
CN52
M11	
M13
M40
M54
M59
CN5
CN35
CN56
M20
M23
M26	
CN23
M2	
M36	
M66	
CN6	
CN7	
CN15	
CN34	
AE003683	G	?	?	?	?	?	?	A	?	?	?	?	?	?	G	C	G	d12	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-		
cDNA	?	?	?	?	?	?	?	?	?	?	?	?	?	?	G	C	C	d12	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
<i>D. simulans</i>	?	?	?	?	?	?	?	?	?	?	?	?	?	?	A	G

Figura 1.2. Polimorfismo nucleotídico en el gen *Fst* en las 30 líneas secuenciadas de las poblaciones de Montemayor (M) y Sant Sadurni d'Anoia (CN). También se incluyen las secuencias del cDNA del gen *Fst* (Goto 2001) y la identificada en el proyecto de secuenciación del genoma de *D. melanogaster* (AE003683). Las posiciones polimórficas están numeradas según el alineamiento múltiple de las secuencias de *D. melanogaster*. Las posiciones polimórficas en la región codificadora se indican sombreadas en gris. NS, polimorfismo no sinónimo. Los puntos (.) indican nucleótidos idénticos a los de la secuencia de referencia y los guiones (-) posiciones con *gaps*. Las deleciones e inserciones de nucleótidos respecto a la secuencia de referencia así como su longitud se indican con d# e i#, respectivamente. El signo de interrogación (?) indica posiciones no analizadas. La última fila muestra la información en *D. simulans* para las posiciones identificadas como polimórficas en *D. melanogaster*.

La región codificadora presenta un total de 16 posiciones polimórficas por cambio nucleotídico. En una de las posiciones se detectaron 3 variantes por lo que el número mínimo de mutaciones es 17. De estas 17 mutaciones, 8 son sinónimas y 9 no sinónimas. Doce posiciones polimórficas presentan variantes únicas. Además de los polimorfismos nucleotídicos, la región codificadora del gen *Fst* presenta 4 polimorfismos por inserción/delección (dos de 24 nucleótidos, uno de 9 nucleótidos y uno de 48 nucleótidos). En las posiciones afectadas por *gaps* se identificaron asimismo dos posiciones polimórficas adicionales con cambios nucleotídicos sinónimos.

Análisis del polimorfismo aminoacídico

En la figura I.3 se muestra el alineamiento múltiple de la secuencia aminoacídica de la proteína codificada por el gen *Fst* deducida a partir de la secuencia nucleotídica de las 30 líneas de *D. melanogaster* analizadas en este estudio. El alineamiento completo comprende un total de 306 posiciones de aminoácidos incluyendo las posiciones con inserción/delección. Se detectaron nueve polimorfismos de reemplazamiento, de los que 6 presentan variantes únicas. Entre los polimorfismos detectados se encontraron cambios aminoacídicos tanto conservativos como no conservativos. Por ejemplo, los cambios fenilalanina/isoleucina (Phe/Ile) y arginina/lisina (Arg/Lys) en las posiciones 543 y 586, respectivamente, no implican cambio ni de carga ni de polaridad y por lo tanto puede considerarse conservativos. Por el contrario, los cambios serina/fenilalanina (Ser/Phe) y metionina/lisina (Met/Lys) detectados en las posiciones 664 y 586, respectivamente, pueden considerarse radicales ya que implican un cambio de polaridad. Al considerar la secuencia de *D. simulans*, se detectaron dos posiciones polimórficas con variantes derivadas casi fijadas en *D. melanogaster* y que corresponden a la arginina en posición 586 y la serina en posición 917. También son derivadas la asparraguina que segrega en la posición 1068 y la lisina de la posición 1125, ambas presentes en las mismas tres líneas. Los restantes cambios derivados corresponden a variantes únicas (figura I.3).

Además de los polimorfismos por reemplazamiento, también se detectaron polimorfismos por inserción/delección en la región codificadora. Ninguno de estos polimorfismos provoca desplazamientos de la pauta de lectura ya que todos ellos provocan inserciones o deleciones con una longitud que es múltiplo de 3. Estas inserciones/deleciones afectan o bien a la región con varios residuos de prolinas

localizada alrededor de la posición 120 de la proteína o bien a la región con repeticiones PEEST localizada hacia el extremo carboxi-terminal de la proteína. Así, en la línea CN1 se detectó una delección de 3 prolina (PPP) con respecto a la secuencia de referencia (figura I.3). Por lo tanto, esta línea presenta 7 prolina consecutivas mientras que este número es de 10 en las restantes líneas. Por otro lado, se detectaron dos delecciones independientes de los motivos PEESTSEA y SEAPEEST en las líneas CN1 y CN34, respectivamente. La delección presente en la línea CN34 se había previamente identificado en la secuencia del cDNA (Goto 2001). Finalmente, en la línea CN15 se identificó una inserción de 16 aminoácidos (TEAPVESTSEAPEEST) que incluye dos motivos PEEST aunque uno de ellos es ligeramente variante.

	543	586	604	627	639	664	736	783	917	1054	1068	1125	1140	1143	1204	1250	1311
M47	Phe	Arg	Gln	Gln	Asn	Ser	Asn	Pro	Ser	Gly	Asp	Glu	Ser	Thr	Pro	-	Thr
M55
CN1	d3	d8
CN9
CN16
CN18
CN22
CN25
CN26
CN41
CN52
M11
M13
M40
M54
M59
CN5
CN35
CN56
M20	Asn	Lys
M23	Asn	Lys
M26	Asn	Lys
CN23	Asn	Lys
M2	.	Met
M36	.	.	Arg
M66	Phe
CN6	.	Lys
CN7	Ileu	Met
CN15	.	.	.	Lys	i16	.
CN34	Thr	d8
AE003683	Asp	.	Ser	d4	Ser	.	.	.
cDNA	Asp	.	.	d4	.	Asp	Leu	.	d8
<i>D. simulans</i>	.	Lys	d1	Thr

Figura I.3. Polimorfismo aminoacídico de la proteína codificada por el gen *Fst* en las 30 líneas secuenciadas de las poblaciones de Montemayor (M) y Sant Sadurní d'Anoia (CN). Para detalles ver el pie de la figura I.2.

Diversidad haplotípica

Las 30 líneas de *D. melanogaster* estudiadas pueden agruparse en un total de 18 haplotipos diferentes al considerar los polimorfismos detectados por cambio nucleotídico en la región analizada. El número de haplotipos identificados no se ve afectado si se consideran también los polimorfismos por inserción/delección. De los 18 haplotipos identificados, trece corresponden a líneas únicas, tres de ellos están presentes en 2 líneas, uno en 4 líneas y el restante es común a 7 líneas. La diversidad haplotípica es $Hd = 0.9310$.

Al considerar únicamente las posiciones polimórficas detectadas en la región codificadora del gen *Fst*, las 30 líneas analizadas se agrupan en un total de 11 haplotipos que representan 9 variantes proteicas (tabla I.1).

Tabla I.1. Haplotipos de la región codificadora del gen *Fst* en *D. melanogaster*

Haplotipo	Nucleótidos polimórficos															Variante proteica	N	Aminoácidos polimórficos																									
	543	586	604	627	632	664	689	785	872	916	917	986	1088	1125	1154			1175	1241	1340	543	586	604	627	664	916	1088	1125															
	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS															
H1	T	G	A	C	T	C	A	A	G	G	T	G	G	G	A	G	G	C	P1	20	Phe	Arg	Gln	Gln	Ser	Ser	Asp	Glu															
H2	C	C													
H3	C													
H4	C	C	A	A	P2	3	Asn	Lys															
H5	.	T	P3	1	Met														
H6	.	.	G	G	.	.	.	P4	1	.	Arg														
H7	T	P5	1	.	.	Arg	.	Phe														
H8	.	A	.	.	.	C	C	P6	1	.	Lys														
H9	A	T	P7	1	Ileu	Met														
H10	.	.	.	A	C	C	A	.	P8	1	.	.	Lys														
H11	C	.	T	C	C	.	.	.	G	A	.	A	.	P9	1	Thr														

NS, polimorfismos nucleotídicos de reemplazamiento. Los números en la parte superior se refieren a las posiciones polimórficas dentro de la región codificadora de acuerdo a la figura I.2. Los puntos (.) indican nucleótidos o aminoácidos idénticos a la secuencia de referencia. H, haplotipo, N, número de líneas para cada variante proteica. **H1** (M47, M55, CN1, CN9, CN16, CN18, CN22, CN25, CN26, CN41, CN52), **H2** (M11, M13, M40, M54, M59, CN5, CN35, CN56), **H3** (M20), **H4** (M23, M26, CN23), **H5** (M2), **H6** (M36), **H7** (M66), **H8** (CN6), **H9** (CN7), **H10** (CN15), y **H11** (CN34).

La red de haplotipos basada en las sustituciones aminoacídicas de las nueve variantes proteicas presentes en esta población se muestra en la figura I.4. En esta figura, el tamaño de los círculos que representan a cada una de las variantes proteicas es proporcional a la frecuencia de cada variante en la muestra analizada. Como se observa en esta red, existe una variante con una frecuencia alta (P1 está presente en el 66.6% de las líneas analizadas) y ocho variantes (de P2 a P9) con una frecuencia igual o inferior al 10%. De hecho, una de estas variantes (P2) se encontró en tres líneas mientras que las restantes están presentes solamente en una línea. La mayoría de las variantes se

diferencian por una única sustitución aminoacídica, sin embargo la rama que conecta la variante P2 contiene dos sustituciones.

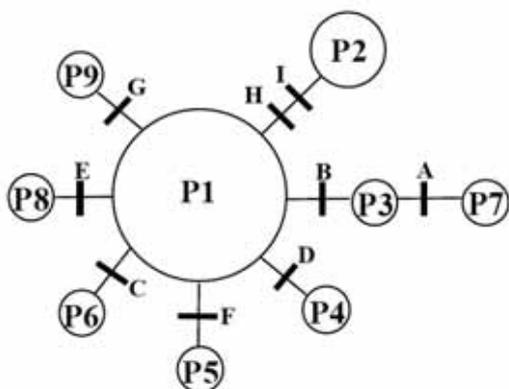


Figura I.4. Red de haplotipos de las 9 variantes de la proteína FROST en *D. melanogaster*. Las marcas y los números en cada rama representan los sitios de reemplazamiento de la siguiente forma: **A**, 543; **B**, 586; **C**, 586; **D**, 604; **E**, 627; **F**, 664; **G**, 916; **H**, 1068; **I**, 1125. El tamaño de los círculos es proporcional a la frecuencia de cada variante.

Desequilibrio de ligamiento y recombinación

El grado de desequilibrio de ligamiento de la región estudiada se analizó para las posiciones polimórficas informativas. De las 21 comparaciones entre los 7 polimorfismos informativos detectados, 4 son significativas por el test χ^2 . Esto representa que un 19% de las comparaciones son significativas por este test. Además, 2 de estas comparaciones permanecen significativas por el test χ^2 después de aplicar la corrección de Bonferroni para múltiples tests. El número mínimo de eventos de recombinación (R_m) en la muestra de las secuencias analizadas se obtuvo a partir del test de los cuatro gametos propuesto por Hudson y Kaplan (1985). Se detectaron únicamente 2 eventos de recombinación en la muestra de secuencias analizadas: uno entre las posiciones 28 y 689 y otro entre las posiciones 986 y 1399.

Diversidad nucleotídica

Los niveles de polimorfismo nucleotídico se estimaron para distintos tipos de posiciones (tabla I.2). La diversidad nucleotídica (π) en el total de la región analizada es 0.0022. Esta estima es similar a la de 0.0028 obtenida al considerar únicamente las posiciones silenciosas (regiones 5' y 3' flanqueantes, y posiciones sinónimas de la región codificadora). Dentro de la región codificadora la estima para las posiciones sinónimas (π_s) es 0.0075 mientras que para las posiciones no sinónimas (π_a) es 0.0014.

De igual forma, para la región no codificadora (regiones 5' y 3' flanqueantes) la estima es 0.0014. Estos resultados indicarían que la selección purificadora actuaría con una intensidad equivalente contra las mutaciones no sinónimas y las producidas en las regiones flanqueantes, mientras que sería más débil al actuar contra las mutaciones sinónimas. Las estimas de la diversidad nucleotídica fueron inferiores a las estimas de θ para todos los tipos de posiciones. Debido a que las estimas de θ dependen únicamente del número de sitios segregantes mientras que las de π están también afectadas por la frecuencia de las diferentes variantes, este resultado indica un exceso de polimorfismos con variantes segregando a bajas frecuencias en la región analizada.

Tabla I.2. Estimaciones del polimorfismo nucleotídico en el gen *Fst* de *D. melanogaster*

Región	Nº de Sitios	S (η)	π	θ
No Codificadora	654	8	0.0014	0.0031
Codificadora	813	16 (17)	0.0028	0.0052
Silenciosa	842.86	16	0.0028	0.0047
Sinónima	188.86	8	0.0075	0.0106
No Sinónima	624.14	8 (9)	0.0014	0.0036
Total	1467	24 (25)	0.0022	0.0041

η , número de mutaciones

Test del neutralismo

Los test del neutralismo propuestos por Tajima (1989) y Fu y Li (1993) se aplicaron para determinar si los patrones de polimorfismo observados en la región estudiada se ajustaban a las predicciones del neutralismo (tabla I.3). Estos tests se realizaron independientemente para los polimorfismos silenciosos y los no sinónimos. Los estadísticos propuestos para aplicar los distintos tests fueron siempre negativos. Todos los tests realizados son significativos ($P < 0.05$) excepto el test de Tajima aplicado al polimorfismo silencioso que es marginalmente significativo ($0.05 < P < 0.1$). Estos resultados indican que el patrón del polimorfismo en la región estudiada no se ajusta a las predicciones del neutralismo. La desviación detectada es debida a un exceso de polimorfismos con variantes únicas en la región estudiada.

Tabla I.3. Tests del neutralismo en el gen *Fst* de *D. melanogaster*

Test	Silenciosa	No sinónima	Total
Tajima D	- 1.3993	- 1.8812*	- 1.7146*
<i>P(D < D_{obs})</i>	0.064	0.011	0.024
Fu y Li D*	- 2.5524*	- 2.1238*	- 2.7070*
<i>P(D* < D*_{obs})</i>	0.023	0.021	0.012
Fu y Li F*	- 2.5695*	- 2.3953*	- 2.810*
<i>P(F* < F*_{obs})</i>	0.021	0.021	0.020
Fay y Wu	- 0.8183	- 5.0758**	- 4.0276
<i>P(H < H_{obs})</i>	0.186	0.008	0.107

Probabilidad de obtener un valor del test estadístico menor que el observado tras realizar simulaciones de coalescencia sin recombinación.

* $0.01 < P < 0.05$, ** $P < 0.01$

El test H de Fay y Wu (2000) fue empleado para poder determinar si existe un exceso de mutaciones derivadas segregando a alta frecuencia en la región analizada. El estadístico H se estimó tanto para toda la región estudiada como independientemente para las posiciones de reemplazamiento y las silenciosas (tabla I.3). La secuencia de *D. simulans* se empleó como referencia para poder determinar el estado de los cambios observados. En todos los casos este test es negativo, sin embargo el test es estadísticamente significativo únicamente para las posiciones de reemplazamiento. Este resultado indica que en la región codificadora del gen *Fst* en *D. melanogaster* existe un exceso de mutaciones derivadas de reemplazamiento que segregan a alta frecuencia.

El test de Mann-Whitney (1999) puede ser utilizado para comparar la distribución de la frecuencia de las variantes polimórficas entre dos categorías de cambios. En este caso, se comparó el espectro de frecuencias del polimorfismo silencioso y de reemplazamiento observado en la región estudiada (figura I.5). Los resultados obtenidos no indican que estos dos tipos de polimorfismos estén segregando a diferentes frecuencias ($z = 0.808$; $P = 0.418$).

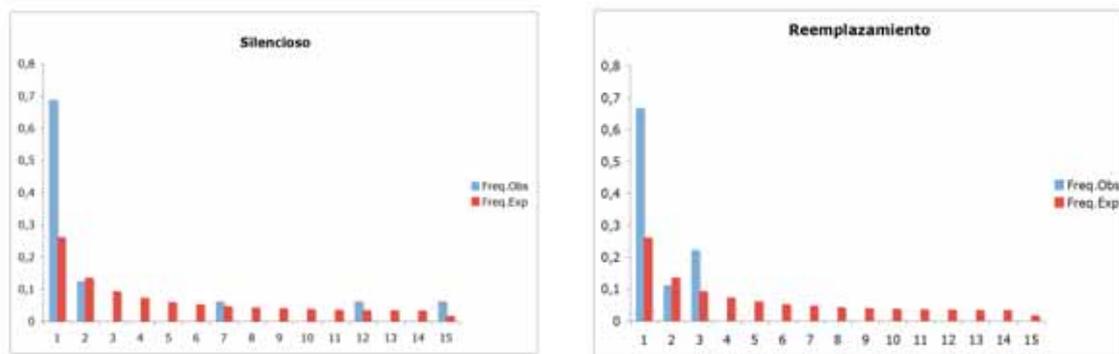


Figura I.5. Espectro de frecuencias del polimorfismo silencioso y de reemplazamiento en el gen *Fst* de *D. melanogaster*

El test HKA (Hudson, Kreitman y Aguadé 1987) se utilizó para contrastar si los niveles de polimorfismo silencioso en *D. melanogaster* y de divergencia silenciosa entre *D. melanogaster* y *D. simulans* detectados en la región *Fst* diferían significativamente de los presentes en la región 5' del gen *Adh* elegida como región control que presumiblemente evoluciona de acuerdo al neutralismo (Hudson, Kreitman y Aguadé 1987). No se detectaron diferencias significativas en los niveles de polimorfismo y divergencia silenciosa de ambas regiones ($\chi^2 = 0.170, P = 0.679$).

También se aplicó el test de MacDonald y Kreitman (1991) para contrastar si la relación entre el número de polimorfismos dentro de especie y el número de diferencias fijadas entre especies es equivalente para los cambios sinónimos y no sinónimos. Al aplicar este test también se utilizó *D. simulans* en la comparación interespecífica. No se detectaron diferencias significativas entre ambas relaciones (tabla I.4). No obstante, el índice de neutralidad (NI) propuesto por Rand y Kann (1996) es mayor que 1 tal y como ocurre cuando hay un exceso de polimorfismos no sinónimos.

Tabla I.4. Test de MK (1991) en el gen *Fst* de *D. melanogaster*

	<u>Polimorfismo en</u> <u><i>D. melanogaster</i></u>	<u>Diferencias fijadas</u> <u>(<i>D. melanogaster</i>/<i>D. simulans</i>)</u>
No sinónimo	9	17
Sinónimo	8	22
		G value = 0.416; P = 0.519
		NI = 1.46

Genealogía de los alelos

La figura I.6 muestra la genealogía de los diferentes alelos de las poblaciones de Montemayor y de Sant Sadurní d'Anoia incluyendo también la línea analizada en el proyecto de secuenciación del genoma de *D. melanogaster* (AE003683). La secuencia del cDNA de Goto (2001) no fue incluida en este análisis debido a que la longitud de la región secuenciada en esta línea es muy inferior a la de las líneas restantes. En este análisis se utilizó *D. simulans* como especie *outgroup*. En el árbol se pueden identificar claramente las 7 líneas que comparten el mismo haplotipo (M47, M55, CN9, CN18, CN25, CN41 y CN52), las 4 que comparten otro haplotipo (M40, M54, CN5 y CN35), y los tres casos de haplotipos compartidos por 2 líneas (CN16-CN22, M23-M26 y M11-M59). Además, no se observan agrupaciones de alelos por localidad, lo que concuerda con el hecho de que las poblaciones analizadas no se encuentran genéticamente diferenciadas.

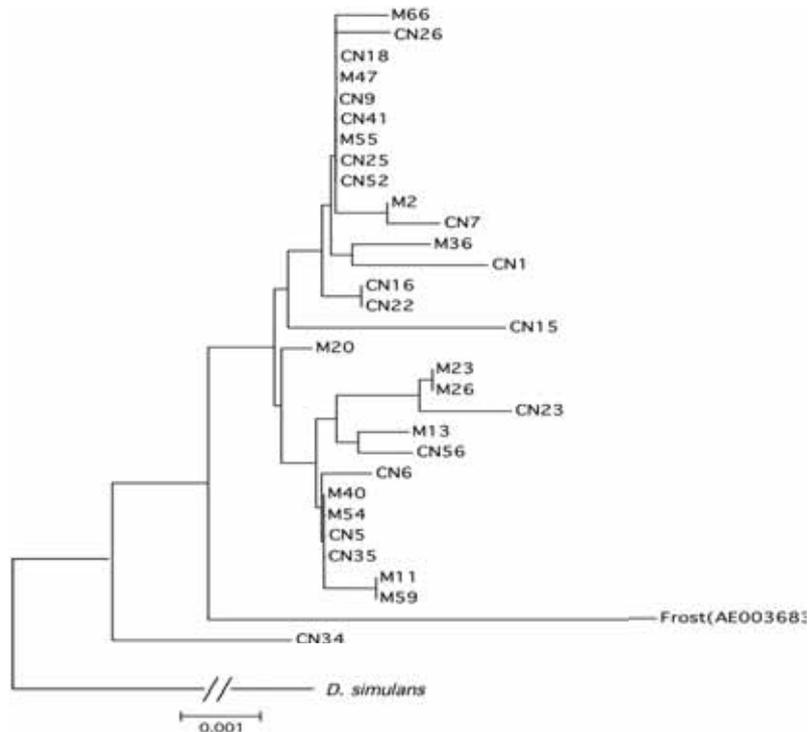


Figura I.6. Genealogía de los alelos de las poblaciones de Montemayor (M) y Sant Sadurní d'Anoia (CN) reconstruida por el método de *neighbor-joining* (Saitou y Nei 1997) usando la distancia de Kimura de 2 parámetros (Kimura 1980).

I.1.1.2 POLIMORFISMO NUCLEOTÍDICO DEL GEN *Fst* EN *D. SUBOBSCURA*

Se secuenció una región de aproximadamente 1.7 kb en 14 líneas de *D. subobscura* de una población natural de Galicia (España). El fragmento secuenciado incluye la región codificadora del gen *Fst* que comprende 675 pb en la gran mayoría de las líneas y parte de las regiones flanqueantes 5' y 3' con una longitud aproximada de 0.6 kb y 0.5 kb, respectivamente.

Ubicación citológica del gen *Fst*

Los resultados de la hibridación *in situ* del gen *Fst* en los cromosomas politénicos de *D. subobscura* muestra una única señal ubicada en la banda citológica 7A del cromosoma O a una sección cromosómica de uno de los puntos de ruptura de la inversión 7, concretamente el más cercano al centrómero (figura I.7).



Figura I.7. Hibridación *in situ* del gen *Fst* en los cromosomas politénicos de *D. subobscura*. La flecha indica la señal de hibridación.

Análisis del polimorfismo nucleotídico

El alineamiento múltiple de las 14 líneas de *D. subobscura* presenta un total de 1788 posiciones. Este número se reduce a 1626 al eliminar las posiciones que presentan *gaps* en una o varias de las líneas analizadas. En el alineamiento se detectaron tanto polimorfismos por cambio nucleotídico como polimorfismos por inserción/delección. Se

Doce de los 15 polimorfismos por inserción/delección identificados se localizan en la región 5' flanqueante. En esta región también se detectaron un total de 24 polimorfismos por cambio nucleotídico, de los que 9 presentan variantes únicas. En una de las posiciones polimórficas segregaban 3 variantes por lo que el número mínimo de mutaciones en esta región es 25. En la región codificadora se identificaron un total 24 posiciones polimórficas por cambio nucleotídico de las que 13 presentan variantes únicas. Trece de los 24 polimorfismos de la región codificadora son sinónimos y 11 no sinónimos. Además de los polimorfismos por cambio nucleotídico, la región codificadora presenta un polimorfismo por inserción/delección de 90 pb. La región 3' flanqueante presenta un total de 31 polimorfismos por cambio nucleotídico y dos por inserción/delección. De los polimorfismos por cambio nucleotídico 15 presentan variantes únicas (Figura I.8).

Análisis del polimorfismo aminoacídico

El alineamiento múltiple de la secuencia aminoacídica de la proteína codificada por el gen *Fst* en las 14 líneas de *D. subobscura* analizadas se muestra en la figura I.9. Este alineamiento comprende un total de 224 posiciones aminoacídicas. Se hallaron un total de 11 polimorfismos de reemplazamiento de los cuales 8 presentan variantes únicas.

		620	642	766	794	880	972	1030	1048	1151	1167	1245	1258
AU89.784	P1	Leu	Ala	Ans	Thr	Asp	Gly	Glu	-	Gly	Ser	Gly	Glu
ES89.484	P2	.	Val
LS89.517	P3	.	Val	Asp	-
LS89.669		.	Val	Asp
S89.174	P4	.	Val	.	Ala	.	.	.	90
AU89.898	P5	.	Val	.	Ala	.	.	Asp	-
AU89.986		.	Val	.	Ala	.	.	Asp	-
LS89.567	P6	.	Val	Asp	-	Arg	.	.	.
S89.4	P7	.	Val	.	.	.	Ala	Asp	-
LS89.754	P8	.	Val	.	Ala	.	.	Asp	-	.	Asn	.	.
AU89.894	P9	.	Val	Lys	Ala	.	.	Asp	-	.	.	.	Asp
AU89.817	P10	.	.	.	Ala	.	.	Asp	-	.	.	Ala	.
AU89.1002	P11	.	.	.	Ala	Glu	.	Asp	-
AU89.918	P12	Met	.	.	Ala	.	.	.	-
<i>D. guanche</i>		.	Val	Asp	-
<i>D. madeirensis</i>		.	Val	Asp	-

Figura I.9. Polimorfismo aminoacídico de la proteína codificada por el gen *Fst* en las 14 líneas de *D. subobscura* secuenciadas de la población de Galicia. Para detalles ver el pie de la figura I.8.

Los polimorfismos aminoacídicos pueden ser clasificados como conservativos o radicales dependiendo de si implican o no un cambio en alguna propiedad físico-química de los aminoácidos. La mayoría de los polimorfismos detectados no implican cambio en la polaridad del aminoácido y por lo tanto pueden considerarse conservativos respecto a esta propiedad. Los polimorfismos leucina/metionina (Leu/Met) y alanina/valina (Ala/Val) en las posiciones 620 y 642, respectivamente, son dos ejemplos. Sin embargo, el polimorfismo Thr/Ala en la posición 794 puede considerarse radical ya que implica un cambio de polaridad. La mayoría de polimorfismos aminoacídicos identificados también pueden considerarse conservativos con respecto a la carga, como por ejemplo los cambios ácido aspártico/ácido glutámico (Asp/Glu) y serina/asparagina (Ser/Asn) en las posiciones 880 y 1167, respectivamente. Sin embargo, el polimorfismo Ans/Lys en la posición 1151 sí que implica un cambio de carga y por lo tanto puede considerarse radical. Al considerar la secuencia de *D. guanche*, todos los aminoácidos que están segregando a baja frecuencia en las posiciones polimórficas son derivados.

En la región codificadora, además de los polimorfismos por reemplazamiento también se detectó un polimorfismo por inserción/delección. Con respecto a la secuencia de referencia, este polimorfismo es una inserción de 30 aminoácidos localizada hacia el extremo carboxi-terminal de la proteína en la línea S89.174. A diferencia de *D. melanogaster* donde se detectaron inserciones/delecciones que implicaban un cambio en el número de motivos PEEST, la inserción en *D. subobscura* no incluye ninguno de estos motivos.

Diversidad haplotípica

Al considerar los polimorfismos detectados por cambio nucleotídico de toda la región analizada, las 14 líneas de *D. subobscura* presentan un total de 14 haplotipos diferentes. Es decir, cada línea corresponde a un haplotipo. Por lo tanto, la diversidad haplotípica es 1.

Los 14 haplotipos de la región codificadora representan un total de 12 variantes proteicas. Como se observa en la figura I.9, existen dos variantes proteicas (P3 y P5) que están presentes en dos de las líneas analizadas (14.28%) y se diferencian por un cambio aminoacídico en la posición 794. El resto de variantes (P1, P2, P4, P6-P12) están presentes en una sola línea.

Desequilibrio de ligamiento y recombinación

El grado de desequilibrio de ligamiento de la región estudiada se analizó para las posiciones polimórficas informativas. De las 861 comparaciones entre los 42 polimorfismos informativos detectados, 46 son significativos por el test χ^2 . Esto representa que un 5.3% de las comparaciones son significativas de acuerdo a este test. El número mínimo de eventos de recombinación (R_m) en la muestra de secuencias analizadas se obtuvo a partir del test de los cuatro gametos propuesto por Hudson y Kaplan (1985). Se detectaron un total de 23 eventos de recombinación en la región analizada.

Diversidad nucleotídica

Los niveles del polimorfismo nucleotídico se estimaron para distintos tipos de posiciones (tabla I.5). La diversidad nucleotídica silenciosa (π_{sil}) es 0.0173 y bastante superior a la detectada en *D. melanogaster* ($\pi_{sil} = 0.0028$, tabla I.2). Dentro de la región codificadora, la estima para las posiciones sinónimas (π_s) es 0.0317 mientras que para las posiciones no sinónimas (π_a) es 0.0057. En la región no codificadora (regiones 5' y 3' flanqueantes) la estima de 0.0154 es considerablemente inferior a la diversidad nucleotídica sinónima ($\pi_s = 0.0317$). Las estimas de θ para la mayoría de posiciones fueron más altas que las estimas de la diversidad nucleotídica. Este resultado indica un exceso de polimorfismos con variantes segregando a bajas frecuencias en la región analizada.

Tabla I.5. Estimaciones del polimorfismo nucleotídico en el gen *Fst* de *D. subobscura*

Región	Nº de Sitios	S (η)	π	θ
No Codificadora	1041	57	0.0154	0.0172
Codificadora	585	24	0.0116	0.0129
Silenciosa	1176.5	70	0.0173	0.0187
Sinónima	135.5	13	0.0317	0.0301
No sinónima	446.512	11	0.0057	0.0077
Total	1626	80 (81)	0.0141	0.0156

η , número de mutaciones

Test del neutralismo

Los test propuestos por Tajima (1989) y Fu y Li (1993) fueron utilizados para detectar posibles desviaciones del modelo neutro en el patrón del polimorfismo observado en la región *Fst* (tabla I.6). Estos test fueron realizados de manera independiente para las posiciones no sinónimas de la región codificadora y para las posiciones silenciosas. Los estadísticos estimados para aplicar los distintos test son negativos en todos los casos, no obstante, ninguno fue significativo. Por lo tanto, los resultados obtenidos indican que el patrón del polimorfismo en la región estudiada se ajusta a las predicciones del neutralismo. Sin embargo, existe un ligero exceso de variantes únicas sobre todo en las posiciones no sinónimas en las que los test de Fu y Li son marginalmente significativos.

El test H de Fay y Wu (2000) fue empleado para determinar la posible existencia de un exceso de mutaciones derivadas segregando a alta frecuencia en la región analizada. El estadístico H se estimó tanto para toda la región estudiada como para las posiciones de reemplazamiento y las silenciosas independientemente (tabla I.6). La secuencia de *D. madeirensis* se empleó como referencia para poder determinar el estado de los cambios observados. El test es negativo tanto en toda la región analizada como en las posiciones silenciosas, sin embargo, no es estadísticamente significativo en ninguno de los casos, lo que indica que no existe un exceso de mutaciones derivadas segregando a alta frecuencia en la región analizada.

Tabla I.6. Tests del neutralismo en el gen *Fst* de *D. subobscura*

Test	Silenciosa	No sinónima	Total
Tajima D	- 0.312	- 1.036	- 0.431
<i>P(D < D_{obs})</i>	0.403	0.141	0.375
Fu y li D*	- 0.476	- 1.657	- 0.678
<i>P(D* < D*_{obs})</i>	0.294	0.087	0.237
Fu y Li F*	- 0.495	- 1.705	- 0.701
<i>P(F* < F*_{obs})</i>	0.314	0.079	0.256
Fay y Wu	- 3.692	1.406	- 2.285
<i>P(H < P < H_{obs})</i>	0.231	0.678	0.260

Probabilidad de obtener un valor del test estadístico menor que el observado tras realizar simulaciones de coalescencia sin recombinación.

El test de MacDonald y Kreitman (1991) se empleó para detectar si la relación entre el número de diferencias no sinónimas y sinónimas fijadas entre especies es equivalente a la relación entre el número de polimorfismos no sinónimos y sinónimos detectados dentro de especie. Para realizar la comparación interespecífica se utilizaron *D. guanche* y *D. madeirensis* independientemente. No se detectaron diferencias entre ambas relaciones en ninguna de las dos comparaciones (tabla I.7).

Tabla I.7. Test de MK (1991) en el gen *Fst* de *D. subobscura*

	Polimorfismo en	Diferencias fijadas	
	<i>D. subobscura</i>	<i>D. guanche</i>	<i>D. madeirensis</i>
No sinónimo	11	10	3
Sinónimo	13	10	1
		G value = 0.076	G value = 1.213
		P = 0.783	P = 0.270
		NI = 0.846	NI = 0.282

Genealogía de los alelos

La genealogía obtenida para los diferentes alelos de la población de Galicia de *D. subobscura* se muestra en la figura I.10. La reconstrucción del árbol genealógico se realizó a partir de toda la región analizada. En este análisis se utilizó como *outgroup* la secuencia de *D. madeirensis*. En el árbol se identifican los 14 haplotipos encontrados en esta población.

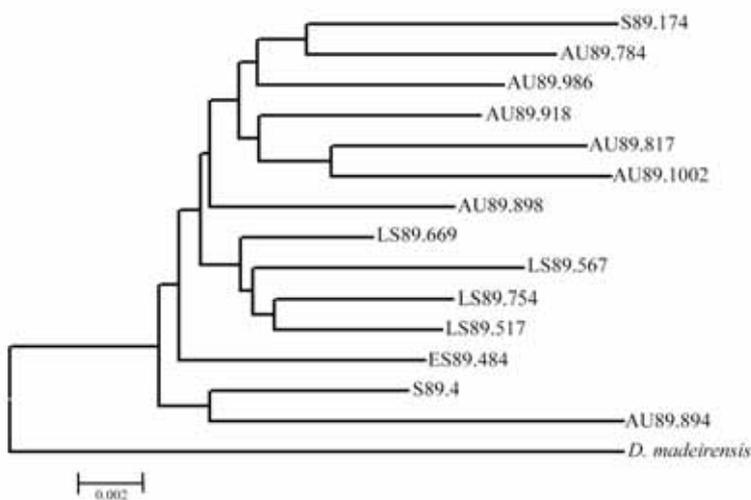


Figura I.10. Genealogía de los 14 alelos analizados en *D. subobscura* reconstruida por el método de *neighbor-joining* (Saitou y Nei 1987) usando la distancia de Kimura de 2 parámetros (Kimura 1980).

I.1.2 DISCUSIÓN

I.1.2.1 DIVERSIDAD NUCLEOTÍDICA DEL GEN *Fst* EN *D. MELANOGASTER* Y *D. SUBOBSCURA*

Los organismos multicelulares responden a los cambios de temperatura por medio de alteraciones en la composición de la membrana celular, la tasa metabólica, el PH intracelular, la concentración de iones y la expresión génica. No obstante, mientras que las respuestas celulares producidas por el estrés causado por el calor han sido bien caracterizadas no ocurre lo mismo en el caso del estrés producido por las bajas temperaturas. Sin embargo, recientemente se han realizado estudios en *Drosophila* con el objetivo de identificar genes que presenten diferencias en los niveles de expresión génica durante la exposición y recuperación del estrés causado por el frío (Goto 2001; Qin et al. 2005; Laayouni et al. 2007; Sinclair et al. 2007), así como análisis de la estructura genética de fenotipos de termo tolerancia en diferentes poblaciones de *D. melanogaster* y *D. subobscura* (Anderson et al. 2003; Morgan y Mackay 2006; Laayouni et al. 2007). Los resultados obtenidos en los anteriores estudios indican que el gen *Fst* es uno de los genes que se ve altamente sobreexpresado en *D. melanogaster* durante la fase de recuperación tras un choque térmico (Goto 2001, Qin et al. 2005; Sinclair et al. 2007) al igual que durante la exposición a un estrés inducido por desecación (Sinclair et al. 2007). El gen *Fst* también se expresa diferencialmente en individuos de poblaciones de *D. subobscura* que han sido expuestos a diferentes regímenes de temperatura (13°C vs 22°C) durante un periodo de tiempo largo (Laayouni et al. 2007). Por otro lado, se ha detectado variación genética significativa para dos fenotipos de termo tolerancia entre dos líneas de *D. melanogaster* (Oregon-R y 2b) e identificado siete QTL (loci que controlan caracteres cuantitativos), tres de los cuales influyen en la resistencia al frío (Morgan y Mackay 2006). Dentro de uno de los anteriores QTL se hallaron tres genes que previamente habían sido relacionados con la respuesta al frío, entre ellos el gen *Fst*. Por lo tanto, ambos tipos de metodología han sugerido que el gen *Fst* se encuentra implicado en la resistencia al estrés causado por el frío en *Drosophila*.

En el presente trabajo se ha estudiado el nivel y patrón del polimorfismo nucleotídico del gen *Fst* en dos poblaciones naturales de *D. melanogaster* ubicadas al sur y al norte de la Península Ibérica: Montemayor (Córdoba) y Sant Sadurní d'Anoia

(Barcelona), respectivamente. A pesar de las diferencias climáticas que presentan estas dos localidades debido a su diferente latitud, no se detectó diferenciación genética en la región *Fst* en las líneas analizadas de *D. melanogaster*. Por lo tanto, si el gen *Fst* tiene un valor adaptativo, es posible que las diferencias climáticas entre las dos poblaciones sean insuficientes para poder determinar una diferenciación genética y provocar un efecto clinal en el patrón de la variabilidad de este gen. La ausencia de diferenciación genética permitió analizar las muestras de ambas poblaciones de *D. melanogaster* conjuntamente y comparar la variabilidad nucleotídica del gen *Fst* en esta especie con la presente en *D. subobscura*.

La diversidad nucleotídica de la región *Fst* es un orden de magnitud superior en *D. subobscura* que en *D. melanogaster* ($\pi_{\text{sil}} = 0.0173$ y $\pi_{\text{sil}} = 0.0028$, respectivamente). Los niveles más altos de variación en el gen *Fst* de *D. subobscura* podrían ser atribuidos a un mayor tamaño efectivo de esta especie en relación a *D. melanogaster*. Sin embargo, si estas diferencias en tamaño efectivo fueran reales deberían manifestarse en otras regiones del genoma. Al comparar la diversidad nucleotídica del gen *Fst* con la presente en otros genes analizados en las mismas especies no se observa este hecho. Aunque los niveles de variación en el gen *Fst* de *D. subobscura* pueden considerarse típicos para la especie, estos niveles en *D. melanogaster* son considerablemente inferiores a los característicos de esta especie. Por lo tanto, las diferencias detectadas en los niveles de variación del gen *Fst* parecen ser particulares de este gen y no pueden explicarse por posibles diferencias en el tamaño efectivo de ambas especies.

La estima de la diversidad nucleotídica silenciosa detectada en el gen *Fst* de *D. subobscura* ($\pi_{\text{sil}} = 0.0173$) presenta un valor relativamente elevado en comparación al observado en otras regiones autosómicas analizadas en esta especie (tabla I.8). Únicamente los genes *Xdh*, *AcpH-1(O₃₊₄)* y *Dca* presentan niveles de polimorfismo silencioso superiores a *Fst*. Sin embargo, la estima de π_{sil} en la región *Fst* es tan solo ligeramente superior a la esta estima promedio ($\pi_{\text{sil}} = 0.0122$) de las 22 regiones estudiadas.

Tabla I.8. Diversidad nucleotídica en diferentes regiones genómicas estudiadas en *D. subobscura*

	Cromosoma/ Ordenación	π_{ul}	π_{total}	Referencia
Autosómicos				
Xdh	O	0.0337	0.0215	Comerón (1997)
<i>Acpi</i> 70A	J	0.0168	0.0140	Cirera y Aguadé (1998)
<i>rp</i> 49	O ₃₊₄	0.0101	0.0080	Rozas et al. (1999)
<i>rp</i> 49	O _{ST}	0.0080	0.0063	Rozas et al. (1999)
<i>Acpi</i> -1	O ₃₊₄	0.0219	0.0123	Navarro-Sabaté et al. (1999)
<i>Acpi</i> -1	O _{ST}	0.0159	0.0097	Navarro-Sabaté et al. (1999)
S25	O _{ST} /O ₃₊₄	0.007/0.010	0.006/0.010	Munté et al. (2005)
P22	O _{ST} /O ₃₊₄	0.005/0.013	0.005/0.011	Munté et al. (2005)
P154	O _{ST} /O ₃₊₄	0.013/0.010	0.005/0.005	Munté et al. (2005)
P2	O _{ST} /O ₃₊₄	0.007/0.009	0.006/0.008	Munté et al. (2005)
S1	O _{ST} /O ₃₊₄	0.004/0.008	0.004/0.007	Munté et al. (2005)
P21	O _{ST} /O ₃₊₄	0.006/0.010	0.004/0.006	Munté et al. (2005)
OS-region	O ₃₊₄	0.0141	-	Sánchez-Gracia (2005)
OS-region	O ₃₊₄₊₂₃	0.0085	-	Sánchez-Gracia (2005)
<i>Fst</i>	O ₃₊₄₊₇	0.0173	0.0141	Presente estudio
<i>Dca</i>	O ₃₊₄₊₇	0.0201	0.0140	Presente estudio
Ligados al sexo				
<i>Pgd</i>	A _{ST}	0.0081 (0.0108)	0.0054 (0.0072)	Martin-Campos (1998)
<i>Rpl</i> 215	A _{ST}	0.0140 (0.0187)	0.0059 (0.0079)	Llopart (1999)
<i>y</i>	A _{ST}	0.0082 (0.0109)	0.0064 (0.0086)	Munté et al. (2000)
<i>y</i>	A ₁	0.0045 (0.0060)	0.0036 (0.0048)	Munté et al. (2000)
P236	A _{ST} /A ₂	0.0076 (0.0101)/0.0060 (0.0080)	0.0076 (0.0101)/0.0060 (0.0080)	Nóbrega (2006)
P150	A _{ST} /A ₂	0.0063 (0.0084)/0.0054 (0.0072)	0.0053 (0.0071)/0.0050 (0.0070)	Nóbrega (2006)
<i>Sxl</i>	A _{ST} /A ₂	0.0063 (0.0084)/0.0047 (0.0062)	0.0056 (0.0075)/0.0042 (0.0056)	Nóbrega (2006)
P125	A _{ST} /A ₂	0.0105 (0.0140)/0.0109 (0.0145)	0.0105 (0.0140)/0.0109 (0.0145)	Nóbrega (2006)
P275	A _{ST} /A ₂	0.0069 (0.0092)/0.0061 (0.0081)	0.0069 (0.0092)/0.0061 (0.0081)	Nóbrega (2006)
<i>RC</i>	A _{ST}	0.0060 (0.0080)	0.0050 (0.0040)	Presente estudio
Promedio				
Total		0.0103 (0.0112)	0.0077 (0.0086)	
Autosómicos		0.0122	0.009	
Ligados al sexo		0.0074 (0.010)	0.0062 (0.0082)	

Entre paréntesis se indican las estimas de π corregidas por el factor 4/3 en las regiones ligadas al cromosoma X.

Las líneas analizadas de *D. subobscura* presentan la ordenación O₃₊₄₊₇. El gen *Fst* en *D. subobscura* está localizado en la banda cromosómica 77A del cromosoma O que no se ve afectada por ninguna de estas inversiones. Sin embargo, la región 77A se encuentra localizada relativamente cerca de uno de los puntos de ruptura de la inversión O₇, concretamente el más cercano al centrómero (77B/C). Aunque la presencia de inversiones puede afectar al nivel y al patrón de polimorfismo nucleotídico en las regiones genómicas asociadas con ellas, no se detecta ningún efecto drástico de la inversión O₇ en la variación nucleotídica del gen *Fst*. Probablemente, la inversión O₇ es lo suficiente antigua para que se haya eliminado el hipotético efecto de barrido selectivo en el gen *Fst* ocasionado por el establecimiento de esta inversión en las poblaciones naturales. Por otra parte, la falta de desequilibrio de ligamiento significativo después de

aplicar la corrección de Bonferroni y el relativamente elevado número mínimo de eventos de recombinación detectados en la región *Fst* en la ordenación O_{3+4+7} indica que la recombinación es importante en los homocariotipos. Este resultado no es sorprendente ya que la frecuencia de esta ordenación en la población analizada (79%) es relativamente elevada (Rozas et al. 1995).

Los niveles de variación nucleotídica detectados en el gen *Fst* de *D. melanogaster* ($\pi_{\text{sil}} = 0.0028$, $\pi_s = 0.0075$) son considerablemente bajos en comparación a los valores de 0.013 (Moriyama y Powell 1996; Andolfatto 2001) que corresponden a la estima promedio de la diversidad nucleotídica sinónima de genes autosómicos en poblaciones no africanas de *D. melanogaster*. Por lo tanto, la diversidad nucleotídica del gen *Fst* en *D. melanogaster* puede considerarse sensiblemente baja en comparación con la diversidad típica de la especie. En *Drosophila*, se ha observado que los niveles del polimorfismo nucleotídico están estrechamente relacionados con la tasa de recombinación, siendo muy bajos en las zonas con poca recombinación (Begun y Aquadro 1992; Aguadé y Langley 1994; Aquadro et al. 1994). El resultado anterior estaría así de acuerdo con la idea que el bajo nivel de variabilidad observado en el gen *Fst* de *D. melanogaster* es debido a su localización en una región con baja recombinación en esta especie. De hecho, el gen *Fst* de *D. melanogaster* se encuentra ubicado en la banda 85E del brazo cromosómico 3R (Goto 2001) cerca del centrómero y considerada como una región con recombinación bastante reducida. Según Kliman y Hey (1993) la tasa de recombinación para esta región es de aproximadamente 0.00093 mientras que dicha tasa en la región central de este brazo cromosómico es de alrededor de 0.003. Además, cerca del gen *Fst* se encuentra el gen *Mtn* para el que también se ha descrito un bajo nivel de polimorfismo en *D. melanogaster* (Lange et al. 1990) en relación a *D. ananassae* (Stephan et al. 1994), especie en la que el gen *Mtn* se encuentra en una zona de recombinación normal. Por lo tanto, el bajo nivel de polimorfismo silencioso y el exceso de polimorfismos únicos detectados en la región *Fst* podrían relacionarse con la localización del gen en una zona de recombinación bastante reducida en *D. melanogaster*.

El relativamente bajo nivel de polimorfismo en el gen *Fst* en *D. melanogaster* también podría explicarse por una baja tasa de mutación neutra en esta región que se traduciría igualmente en una baja divergencia. Aunque el test de HKA (Hudson et al. 1987) no detecta un desajuste entre los niveles de polimorfismo y divergencia en la región *Fst* respecto a la región 5' del gen *Adh*, el bajo nivel de polimorfismo silencioso

de la región *Fst* no se corresponde con una baja divergencia. De hecho, la divergencia sinónima en el gen *Fst* es relativamente elevada en comparación a otros genes estudiados en *D. melanogaster* y *D. simulans* (Begun y Whitley 2000a).

Por lo tanto, el bajo nivel de polimorfismo y el exceso de polimorfismos únicos en *D. melanogaster* reflejados por los valores negativos significativos de los estadísticos de los tests de Tajima y Fu y Li, podrían explicarse por un barrido selectivo o *hitchhiking* en o cerca del gen *Fst*. De hecho, la propagación de una mutación seleccionada positivamente en la población elimina la variación neutra que se encuentra ligada a ella, lo que produce una disminución en los niveles de la variación así como un espectro de frecuencias con un exceso de polimorfismos únicos. La baja recombinación de la región *Fst* determinaría que la zona afectada por un hipotético barrido selectivo en esta región fuera mayor que en una región de recombinación normal, considerando una intensidad de selección equivalente en ambas regiones. Además, el exceso de mutaciones derivadas segregando a frecuencias elevadas que se refleja por el signo negativo del estadístico H de Fay y Wu y que es significativo para las mutaciones no sinónimas, apoyaría la idea de un barrido selectivo en o cerca de la región *Fst*.

Además de los factores selectivos, los factores históricos o demográficos podrían contribuir a los relativamente bajos niveles de variación nucleotídica y al exceso de polimorfismos únicos observados en el gen *Fst* de *D. melanogaster*. Aunque esta especie supuestamente se originó en el Este de África, se ha propagado alrededor del mundo principalmente por los humanos (David y Cappy 1988). Así, las poblaciones estudiadas pueden presentar un cierto efecto fundador que determine que no se encuentren en equilibrio mutación-deriva. No obstante, se espera que los factores demográficos afecten a todo el genoma y no a genes particulares. Para poder discernir entre factores demográficos y selectivos se requiere información de distintas regiones genómicas. En la actualidad se dispone de información acerca de la variabilidad nucleotídica en la población de Montemayor de dos genes autosómicos: el gen *Acp70* (Cirera y Aguadé 1997) y los genes que codifican para proteínas de unión a odorantes (*Odorant-binding proteins*) *OS-E* y *OS-F* (Sánchez-Gracia et al. 2003). Los niveles de variabilidad encontrados en el gen *Acp70* ($\pi_{\text{total}} = 0.016$) son muy superiores a los detectados en el gen *Fst* ($\pi_{\text{total}} = 0.0022$), mientras que los encontrados en la región *OS* ($\pi_{\text{total}} = 0.0018$) presentan una alta similitud. No obstante, la región *OS* también está localizada en una zona de recombinación reducida. Por otro lado, el nivel de variación nucleotídica en la población de Sant Sadurní d'Anoia se ha estimado tras el análisis de

109 fragmentos no codificadores localizados a lo largo del cromosoma X (Orengo y Aguadé 2004). La estima promedio de la diversidad nucleotídica es $\pi = 0.0038$ con un rango de variación que oscila entre 0 y 0.0132. Por lo tanto, el nivel de variación en las regiones no codificadoras del gen *Fst* ($\pi_{NC} = 0.0014$) correspondería a un nivel de variación relativamente bajo en la población. El análisis multilocus realizado por Orengo y Aguadé (2004) en la población de Sant Sadurní d'Anoia indicó que los datos sólo eran compatibles y con una baja probabilidad con un cuello de botella de intensidad intermedia relativamente reciente. Por lo tanto, parece poco probable que se haya producido un evento demográfico muy drástico que afecte al nivel de variabilidad nucleotídica de todo el genoma, por lo menos en la población de Sant Sadurní d'Anoia.

El primer estudio recopilatorio del polimorfismo nucleotídico en *D. melanogaster* fue realizado por Moriyama y Powell (1996). Tras el análisis de 15 genes autosómicos detectaron un total de 59 polimorfismos no sinónimos y 133 polimorfismos sinónimos (relación 0.44). Estudios posteriores muestran relaciones de 0.49 (Andolfatto 2001) y 0.50 (Fay et al. 2002) basadas en el análisis de 17 y 31 genes autosómicos, respectivamente. Más recientemente, Bierne y Eyre-Walker (2004) han analizado la tasa de sustituciones aminoacídicas adaptativas en *D. melanogaster*. Este estudio incluye un total de 57 genes, 41 de ellos autosómicos. En esta última muestra detectan 185 polimorfismos no sinónimos y 512 sinónimos, lo que representa un relación de 0.36. Esta relación difiere significativamente de la relación 1.125 detectada en el gen *Fst* con 9 polimorfismos no sinónimos y 8 polimorfismos sinónimos ($G = 5.29$, $P = 0.02$). Además, cabe mencionar que tan solo en 7 de los 41 genes recopilados por Bierne y Eyre-Walker (2004) el número de polimorfismos no sinónimos supera al de sinónimos.

A pesar que el test de McDonald y Kreitman (1991) no es estadísticamente significativo en ninguna de las dos especies, el índice de neutralidad mayor que 1 detectado en *D. melanogaster* indicaría en esta especie una tendencia en el gen *Fst* hacia un exceso de polimorfismo no sinónimo respecto al esperado bajo el modelo neutro. Este tipo de desviación se ha detectado principalmente en los genes del genoma mitocondrial de *Drosophila* y otros organismos (Nachman et al. 1994; Nachman et al. 1996; Rand y Kann 1996). La detección de un exceso de polimorfismo no sinónimo se ha interpretado generalmente de acuerdo al carácter ligeramente deletéreo de las mutaciones no sinónimas que se acumularían como polimorfismos de vida corta (*short-lived polymorphism*) dentro de la especie, pero que no permanecerían en la población el

tiempo suficiente para llegar a fijarse y así contribuir a la divergencia entre especies. Una de las predicciones de la teoría casi neutra de la evolución molecular (Ohta 1992) es que las mutaciones ligeramente deletéreas no deberían alcanzar frecuencias apreciables en la población a diferencia de las mutaciones neutras que sí podrían hacerlo. Por lo tanto, el espectro de frecuencia del polimorfismo no sinónimo y silencioso proporciona posibles test para comprobar esta predicción. El test de Mann-Whitney (1999) fue empleado para comparar la distribución de frecuencias de las mutaciones silenciosas y las no sinónimas. Sin embargo, este test no detectó diferencias significativas en el rango de la distribución de las frecuencias, lo que indica un comportamiento similar de los dos tipos de mutaciones en el gen *Fst* de *D. melanogaster*. De hecho, aunque se detecta un exceso significativo de mutaciones no sinónimas segregando a baja frecuencia (Tajima $D = -1.881$, $P = 0.011$), también se detecta un exceso similar marginalmente significativo para las mutaciones silenciosas (Tajima $D = -1.399$, $P = 0.063$).

I.2 DIVERGENCIA DEL GEN *Fst*

I.2.1 RESULTADOS

I.2.1.1 DIVERGENCIA NUCLEOTÍDICA DEL GEN *Fst* EN EL GRUPO *MELANOGASTER* DE *DROSOPHILA*

El gen *Frost* (*Fst*) fue secuenciado en cinco especies del grupo *melanogaster* de *Drosophila*. El alineamiento múltiple de estas cinco especies comprende un total de 1590 posiciones. Sin embargo, este número se reduce a 1227 posiciones al excluir las posiciones con *gaps*. La estructura del gen *Fst* consiste en un único exón con una longitud de 873, 1023, 894, 939 y 855 pb en *D. melanogaster*, *D. simulans*, *D. mauritiana*, *D. yakuba* y *D. erecta*, respectivamente.

En la tabla I.9 se muestran las estimas de la divergencia nucleotídica entre las diferentes especies para el gen *Fst*. Dentro de la región codificadora, las estimas de divergencia no sinónima (K_a) fueron inferiores a las estimas de divergencia sinónima (K_s) en todas las comparaciones. Además, las estimas de divergencia sinónima fueron siempre superiores a las correspondientes estimas en las regiones no codificadoras (regiones flanqueantes 5' y 3'). La relación K_a/K_s oscila entre 0.126 y 0.243 con un valor promedio de 0.198.

Tabla I.9. Divergencia nucleotídica del gen *Fst* en el grupo *melanogaster*

Especies	$K_{NC}(5')$	$K_{NC}(3')$	K_s	K_a	K_a / K_s
<i>mel-sim</i>	0.027	0.068	0.144	0.035	0.243
<i>mel-mau</i>	0.023	0.076	0.178	0.04	0.224
<i>mel-yak</i>	0.058	0.151	0.513	0.074	0.144
<i>mel-ere</i>	0.055	0.185	0.538	0.116	0.215
<i>sim-mau</i>	0.017	0.036	0.099	0.019	0.191
<i>sim-yak</i>	0.066	0.141	0.466	0.063	0.126
<i>sim-ere</i>	0.062	0.166	0.499	0.105	0.210
<i>mau-yak</i>	0.062	0.134	0.455	0.07	0.153
<i>mau-ere</i>	0.058	0.176	0.479	0.112	0.233
<i>yak-ere</i>	0.066	0.149	0.305	0.074	0.242

K_{NC} , número de sustituciones por posición en las correspondientes regiones no codificadora; K_s , número de sustituciones sinónimas por posición sinónima; K_a , número de sustituciones no sinónimas por posición no sinónima. *mel*, *D. melanogaster*; *sim*, *D. simulans*; *mau*, *D. mauritiana*; *yak*, *D. yakuba*; *ere*, *D. erecta*.

La figura I.11 muestra los árboles filogenéticos reconstruidos por el método de *neighbor-joining* (Saitou y Nei, 1987) en base a la divergencia nucleotídica de la región codificadora del gen *Fst*. La longitud de las ramas es diferente entre *D. yakuba* y *D. erecta* en el caso de la divergencia total (figura I.11A). No obstante, esta diferencia no se observa en el árbol obtenido a partir de la divergencia sinónima (figura I.11B) mientras que es muy acusada al emplear la divergencia no sinónima (figura I.11C).

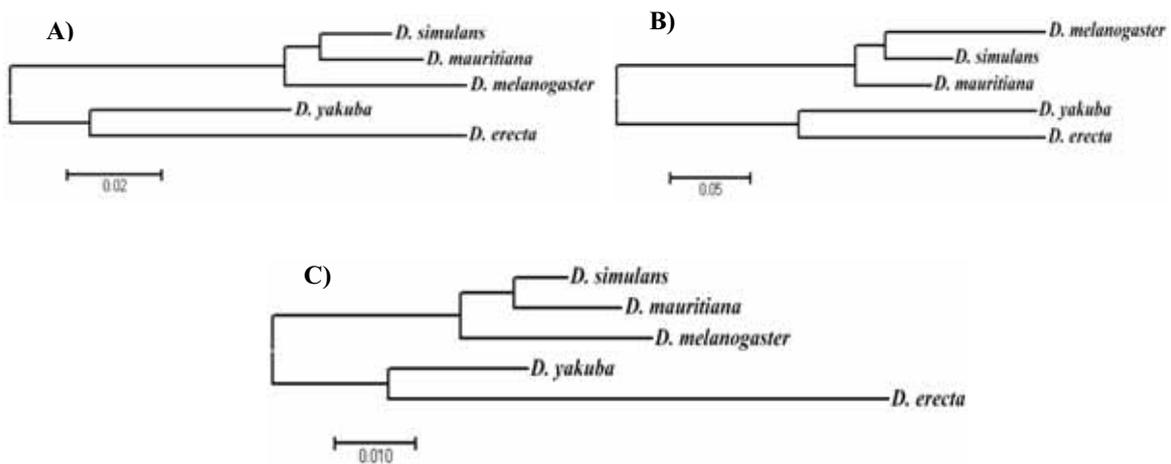


Figura I.11. Árboles filogenéticos obtenidos por el método de *neighbor-joining* (Saitou y Nei 1987) en base a la A) divergencia nucleotídica total de la región codificadora, B) divergencia sinónima (K_s) y C) divergencia no sinónima (K_a) del gen *Fst* en el grupo *melanogaster* de *Drosophila*.

Para poder detectar posibles desviaciones de la hipótesis del reloj molecular se realizó el *relative rate test* propuesto por Wu y Li (1985) en todas las comparaciones posibles de acuerdo a la filogenia obtenida (tabla I.10). Este test no detectó ninguna desviación significativa en la constancia de la tasa de sustitución en ninguna de las comparaciones entre las especies del complejo *melanogaster*. Sin embargo, la diferencia en las tasas de sustitución entre los linajes de *D. yakuba* y *D. erecta* sí que es estadísticamente significativa. Con el objetivo de determinar que tipo de sustitución es la responsable de la desviación detectada en la constancia de la tasa de sustitución nucleotídica en estos linajes se aplicó el test de Tajima (1993). Este test permite contrastar si el número de sustituciones acumuladas en dos linajes difiere o no significativamente. El test se aplicó independientemente para las sustituciones sinónimas y para las no sinónimas producidas en los linajes de *D. yakuba* y *D. erecta*, utilizando *D. melanogaster* como especie de referencia. El resultado del test es altamente significativo para las sustituciones no sinónimas ($\chi^2 = 12.90$, $P < 0.001$) e

indica un exceso de sustituciones no sinónimas acumuladas en el linaje de *D. erecta*. El número de sustituciones sinónimas no difiere significativamente en ambos linajes ($\chi^2 = 0.027$, $P > 0.05$).

Tabla I.10. Relative rate test del gen *Fst* en el grupo *melanogaster*

Out- Group	Especie 1- Especie 2	Test de Wu y Li			
		$K_{13} - K_{23}$	SD	z - score	P value
<i>mel</i>	<i>sim - mau</i>	-0.0080	0.0073	-1.0984	0.275
	<i>yak - ere</i>	-0.0411	0.0144	-2.8434	0.004*
<i>sim</i>	<i>yak - ere</i>	-0.0392	0.0142	-2.7526	0.006*
<i>mau</i>	<i>yak - ere</i>	-0.0356	0.0143	-2.5520	0.01*
<i>yak</i>	<i>mel - sim</i>	0.0176	0.0104	16.957	0.091
	<i>mel - mau</i>	0.0179	0.0111	16.175	0.107
	<i>sim - mau</i>	0	0.0079	0	1
<i>ere</i>	<i>mel - sim</i>	0.0081	0.0108	0.7547	0.453
	<i>mel - mau</i>	0.0131	0.0115	11.235	0.263
	<i>sim - mau</i>	0.0047	0.0083	0.569	0.575

Nombre de las especies como en la tabla I.9.

* $P < 0.05$

También se analizó el sesgo en el uso de codones sinónimos (*codon bias*) en las diferentes especies (tabla I.11). Los niveles de *codon bias* detectados fueron muy similares y relativamente bajos en todas las especies, con una media del valor ENC igual a 49.873. *Drosophila yakuba* fue la especie con el *codon bias* más alto (ENC = 46.499) mientras que *D. erecta* presenta el más bajo (ENC = 52.431).

Tabla I.11. Etimas de *codon bias* del gen *Fst* en el grupo *melanogaster*

Especie	Scaled χ^2	CBI	ENC
<i>D. melanogaster</i>	0.251	0.351	50.937
<i>D. simulans</i>	0.240	0.332	50.264
<i>D. mauritiana</i>	0.253	0.346	49.238
<i>D. yakuba</i>	0.314	0.381	46.499
<i>D. erecta</i>	0.237	0.367	52.431

CBI, índice de *codon bias*; ENC, número efectivo de codones.

I.2.1.1.1 DIVERGENCIA AMINOACÍDICA DE LA PROTEÍNA FROST EN EL GRUPO *MELANOGASTER* DE *DROSOPHILA*

El alineamiento múltiple de la proteína codificada por el gen *Fst* en las 5 especies del subgrupo *melanogaster* comprende un total de 330 aminoácidos (figura I.12). La longitud de la proteína en *D. melanogaster*, *D. simulans*, *D. mauritiana*, *D. yakuba* y *D. erecta* es de 290, 340, 297, 312 y 284 aminoácidos, respectivamente. Estas diferencias en el tamaño se deben principalmente a la variación en el número de repeticiones de los motivos PEEST en la región C-terminal de la proteína. Este número oscila desde 14 en *D. simulans* hasta 4 en *D. erecta*. Además de los motivos PEEST perfectos también se detectaron motivos que difieren en uno de los residuos (PEEIT, PEDST, PEETT, PKEST, TEEEST, AEEST y PVEST). La oscilación en el número de motivos abarca desde 19 en *D. simulans* hasta 11 en *D. erecta* al tener en cuenta estos motivos imperfectos. Los cambios aminoacídicos que presentan los motivos PEDST, PEETT y AEEST son conservativos puesto que no producen ningún cambio en las propiedades físico-químicas de los aminoácidos. Sin embargo, los cambios en los motivos PEEIT, PKEST, TEEEST y PVEST son radicales ya sea para la polaridad, la carga o ambos.

Otra característica a destacar de la proteína FROST es el ser una proteína muy rica en el aminoácido treonina (T). Así, el porcentaje de treonina que presenta la proteína FROST en las especies del grupo *melanogaster* oscila entre 12.47% en *D. melanogaster* hasta 14.41% en *D. simulans*. La mayoría de estos residuos se encuentran en pares separados por un número irregular de aminoácidos en las diferentes especies. El número de pares de treoninas observado es de 12 en *D. melanogaster* (TT_{X6} TT_{X8} TT_{X17} TT_{X11} TT_{X10} TT_{X6} TT_{X10} TT_{X5} TT_{X23} TT_{X6} TT_{X14} TT), 19 en *D. simulans* (TT_{X6} TT_{X8} TT_{X9} TT_{X6} TT_{X9} TT_{X10} TT_{X6} TT_{X10} TT_{X5} TT_{X7} TT_{X6} TT_{X14} TT_{X6} TT_{X13} TT_{X14} TT_{X14} TT_{X14} TT_{X14} TT), 14 en *D. mauritiana* (TT_{X8} TT_{X9} TT_{X6} TT_{X9} TT_{X7} TT_{X6} TT_{X10} TT_{X5} TT_{X7} TT_{X6} TT_{X14} TT_{X29} TT_{X14} TT), 17 en *D. yakuba* (TT_{X6} TT_{X8} TT_{X9} TT_{X6} TT_{X12} TT_{X9} TT_{X6} TT_{X10} TT_{X6} TT_{X11} TT_{X9} TT_{X5} TT_{X13} TT_{X14} TT_{X14} TT_{X14} TT) y 14 en *D. erecta* (TT_{X6} TT_{X12} TT_{X9} TT_{X5} TT_{X12} TT_{X9} TT_{X5} TT_{X13} TT_{X6} TT_{X6} TT_{X6} TT_{X6} TT_{X6} TT).

Otra característica de la proteína codificada por el gen *Fst*, es la presencia de una región rica en prolinas (P). En la figura I.12 esta región se encuentra ubicada entre las posiciones 115 y 124. En *D. melanogaster* esta región incluye un total de 10 prolinas adyacentes (aunque se ha identificado una línea con 7 residuos), mientras que en *D.*

simulans y *D. yakuba* incluye 7 residuos, en *D. erecta* 9 y en *D. mauritiana* uno de los 7 residuos adyacentes presenta un cambio de prolina a alanina.

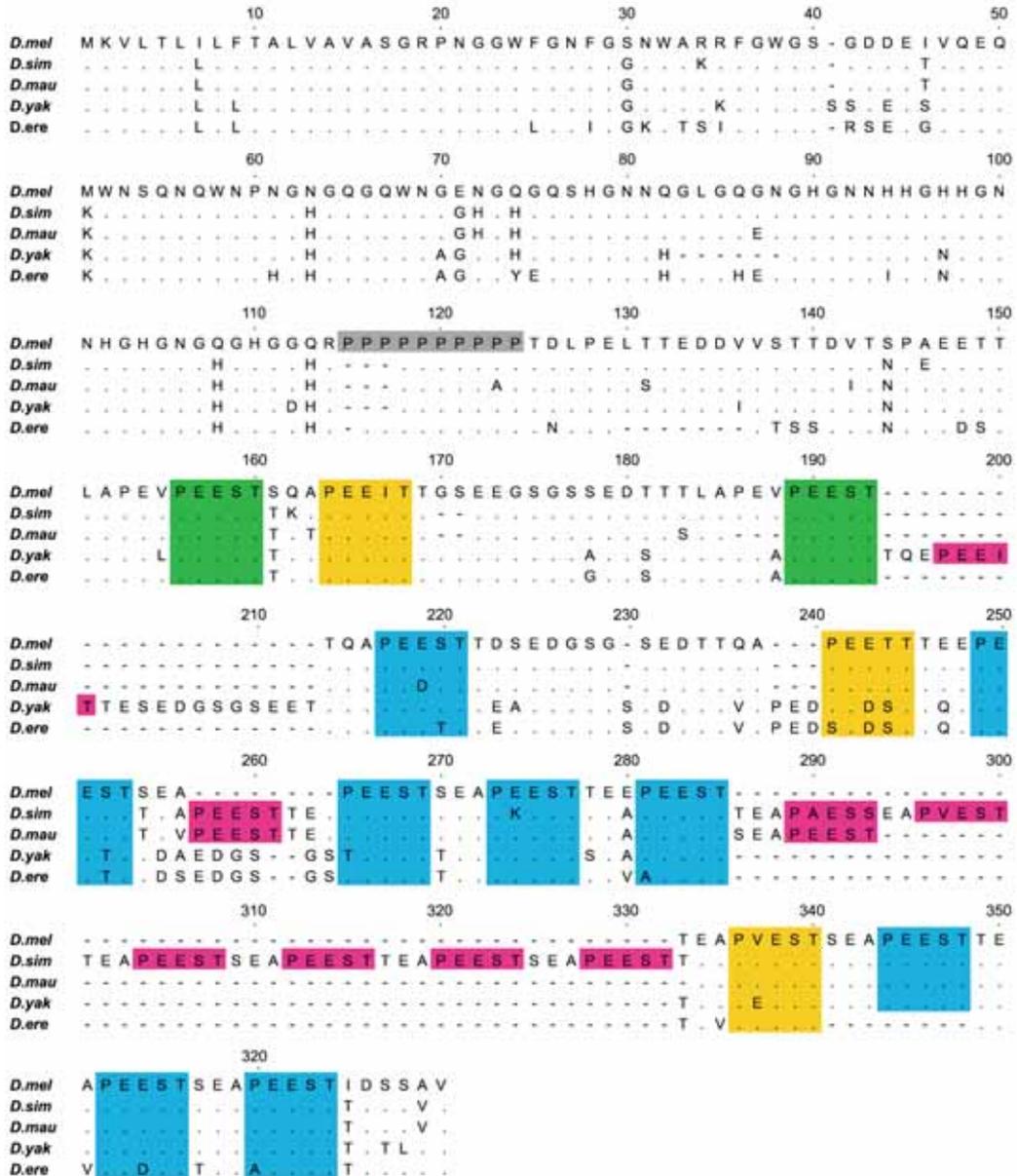


Figura I.12. Alineamiento múltiple de la secuencia aminoacídica de la proteína codificada por el gen *Fst*. Los puntos (.) indican aminoácidos idénticos a los de la secuencia de referencia y los guiones (-) deleciones de aminoácidos respecto a esta secuencia. En gris se indica la región rica en prolina. En verde, los motivos PEEST conservados en todas las especies. En amarillo, los motivos que presentan un cambio en uno de los residuos del motivo PEEST en la secuencia de referencia. En fucsia, las inserciones de motivos con respecto a la secuencia de referencia. En azul, los motivos PEEST presentes en *D. melanogaster* pero que presentan algún cambio en uno de los residuos en alguna o varias de las especies restantes. *D.mel.*, *D. melanogaster*; *D.sim.*, *D. simulans*; *D.mau.*, *D. mauritiana*; *D.yak.*, *D. yakuba* y *D.ere.*, *D. erecta*.

1.2.1.2 DIVERGENCIA NUCLEOTÍDICA DEL GEN *Fst* EN EL GRUPO *OBSCURA* DE *DROSOPHILA*

La región *Fst* fue secuenciada en tres especies del grupo *obscura* de *Drosophila*: *D. subobscura*, *D. madeirensis* y *D. guanche*. La secuencia de *D. pseudoobscura* empleada para realizar los diferentes análisis corresponde a la obtenida en el proyecto de secuenciación del genoma de esta especie. El alineamiento múltiple de las cuatro especies comprende un total de 804 posiciones que se reduce a 561 al eliminar aquellas posiciones que presentan *gaps*. Al igual que en el grupo *melanogaster* de *Drosophila*, la estructura del gen *Fst* en estas cuatro especies consiste en un único exón con una longitud de 783 pb en *D. pseudoobscura* y de 588 pb en *D. subobscura*, *D. madeirensis* y *D. guanche*.

Las estimas de la divergencia nucleotídica detectada para las posiciones no sinónimas (K_a) fueron inferiores a las correspondientes estimas para las posiciones sinónimas (K_s) en todas las comparaciones (tabla I.12). Además, las estimas de divergencia en las regiones no codificadoras (regiones flanqueantes 5' y 3') también fueron siempre inferiores a las estimas de K_s . La relación K_a/K_s oscila entre 0.255 y 0.470 con un valor promedio de 0.325.

Tabla I.12. Divergencia nucleotídica del gen *Fst* en el grupo *obscura*

Especies	$K_{NC(5')}$	$K_{NC(3')}$	K_s	K_a	K_a / K_s
<i>pseu-sub</i>	0.235	0.162	0.684	0.175	0.255
<i>pseu-made</i>	0.238	0.147	0.673	0.182	0.270
<i>pseu-guan</i>	0.247	0.159	0.683	0.19	0.278
<i>sub-made</i>	0.037	0.024	0.024	0.009	0.375
<i>sub-guan</i>	0.069	0.049	0.093	0.027	0.290
<i>made-guan</i>	0.062	0.039	0.068	0.032	0.470

K_{NC} , número de sustituciones por posición en las correspondientes regiones no codificadoras; K_s , número de sustituciones sinónimas por posición sinónima; K_a , número de sustituciones no sinónimas por posición no sinónima. *guan*, *D. guanche*; *pseu*, *D. pseudoobscura*; *mad*, *D. madeirensis*; *sub*, *D. subobscura*.

En la figura I.13 se muestran los árboles filogenéticos obtenidos a partir de la divergencia de la región codificadora del gen *Fst* usando *D. pseudoobscura* como especie *outgroup*. Al analizar la divergencia total y no sinónima se observa que la rama que conduce a *D. guanche* es ligeramente más larga que las que conducen a las especies *D. subobscura* y *D. madeirensis*.

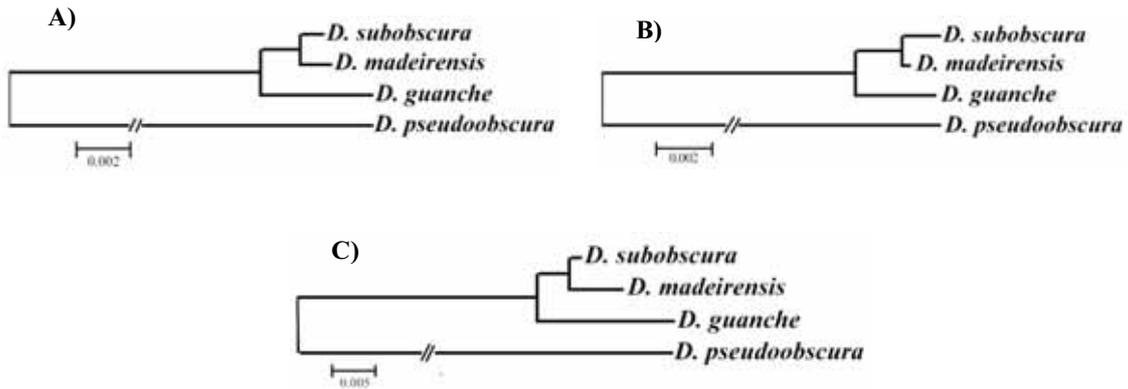


Figura I.13. Árboles filogenéticos obtenidos por el método de *neighbor joining* (Saitou y Nei 1987) en base a la A) divergencia nucleotídica total de la región codificadora, B) divergencia sinónima (K_s) y C) divergencia no sinónima (K_a) del gen *Fst* en el grupo *obscura* de *Drosophila*.

La distribución geográfica de las especies *D. madeirensis* y *D. guanche* está restringida a las islas de Madeira y Canarias, respectivamente. Por el contrario, *D. subobscura* es una especie que presenta una distribución paleártica y, por lo tanto, se espera que su tamaño efectivo (N_e) sea mayor que el de las especies insulares. Para poder determinar si existe un efecto en la tasa de sustitución nucleotídica causado por un N_e menor en *D. madeirensis* y *D. guanche* se aplicó el *relative rate test* (Wu y Li 1985) entre las especies del complejo *subobscura* usando *D. pseudoobscura* como *outgroup* (tabla I.13). En las comparaciones entre *D. subobscura* y *D. guanche* o entre *D. madeirensis* y *D. guanche* usando *D. pseudoobscura* como *outgroup*, los valores negativos de la diferencia $K_{13} - K_{23}$ indican una tasa de sustitución más elevada en el linaje de *D. guanche* (tabla I.13). No obstante, esta diferencia no es estadísticamente significativa. Por otro lado, el *relative rate test* no detectó ninguna desviación significativa de la hipótesis del reloj molecular entre *D. subobscura* y *D. madeirensis* utilizando *D. guanche* (o *D. pseudoobscura*) como especie *outgroup*.

Tabla I.13. Relative rate test del gen *Fst* en el grupo *obscura*

Out- Group	Especie 1- Especie 2	Test de Wu y Li			
		$K_{13} - K_{23}$	SD	z - score	P value
guan.....	sub - made	0.0054	0.0053	1.010	0.317
pseu.....	sub - made	0.0025	0.0065	0.3828	0.764
	sub - guan	-0.0126	0.0107	-1.1762	0.242
	made - guan	-0.0151	0.0101	-1.4836	0.139

Nombre de las especies como en la tabla I.12.

El nivel de *codon bias* que refleja el uso desigual de los codones sinónimos, y que depende entre otras características del tamaño efectivo (N_e), fue estimado para las diferentes especies del grupo *obscura* (tabla I.14). *Drosophila pseudoobscura* fue la especie con el nivel de *codon bias* más elevado (ENC = 46.106). En las especies restantes, estas estimas fueron muy similares y relativamente bajas. La media del valor ENC en las especies del grupo *obscura* es igual a 56.947.

Tabla I.14. Estimaciones de *codon bias* del gen *Fst* en el grupo *obscura*

Especie	Scaled χ^2	CBI	ENC
<i>D. subobscura</i>	0.113	0.261	61
<i>D. madeirensis</i>	0.109	0.261	61
<i>D. guanche</i>	0.153	0.253	59.682
<i>D. pseudoobscura</i>	0.227	0.376	46.106

CBI, índice de *codon bias*; ENC, número efectivo de codones

I.2.1.2.1 DIVERGENCIA AMINOACÍDICA DE LA PROTEÍNA FROST EN EL GRUPO *OBSCURA* DE *DROSOPHILA*

El alineamiento múltiple de la proteína codificada por el gen *Fst* en las cuatro especies del grupo *obscura* analizadas consta de 268 aminoácidos (figura I.14). En las especies *D. subobscura*, *D. madeirensis* y *D. guanche* la longitud de la proteína es de 195 aminoácidos mientras que en *D. pseudoobscura* alcanza un total de 260 aminoácidos. Del mismo modo que lo observado en el grupo *melanogaster* de *Drosophila*, esta diferencia en el tamaño de la proteína se debe principalmente a la variación en el número de repeticiones tanto de los motivos PEEST perfectos como de aquellos que difieren en uno de los residuos (SEEST, PEDST, PVEST, PDEST y TEEST). El número de motivos PEEST oscila entre 1 en *D. subobscura*, *D. madeirensis* y *D. guanche* hasta 3 en *D. pseudoobscura*. Este número aumenta a 6 y 10, respectivamente, cuando se tienen en cuenta los motivos que difieren en uno de los residuos. El porcentaje de treonina que presenta la proteína FROST en este grupo oscila entre 15% en *D. pseudoobscura* hasta 16.92% en *D. subobscura* y *D. madeirensis*. La mayoría de estos residuos se encuentran en pares separados de una forma más regular que la observada en las especies del grupo *melanogaster*. En *D. pseudoobscura* se detectaron 18 pares de treoninas (TT_{X8} TT_{X5} TT_{X6} TT_{X5} TT_{X9} TT_{X9} TT_{X5} TT_{X9} TT_{X9}

TT_{X5} TT_{X9} TT_{X9} TT_{X5} TT_{X9} TT_{X9} TT_{X11} TT_{X9} TT), 9 en *D. subobscura* y *D. madeirensis* (TT_{X9} TT_{X13} TT_{X7} TT_{X9} TT_{X19} TT_{X6} TT_{X29} TT_{X8} TT) y 8 en *D. guanche* (TT_{X9} TT_{X13} TT_{X7} TT_{X9} TT_{X19} TT_{X6} TT_{X39} TT). Por otro lado, ninguna de las especies analizadas del grupo *obscura* presenta la región rica en prolinas detectada en las especies del grupo *melanogaster* de *Drosophila*.

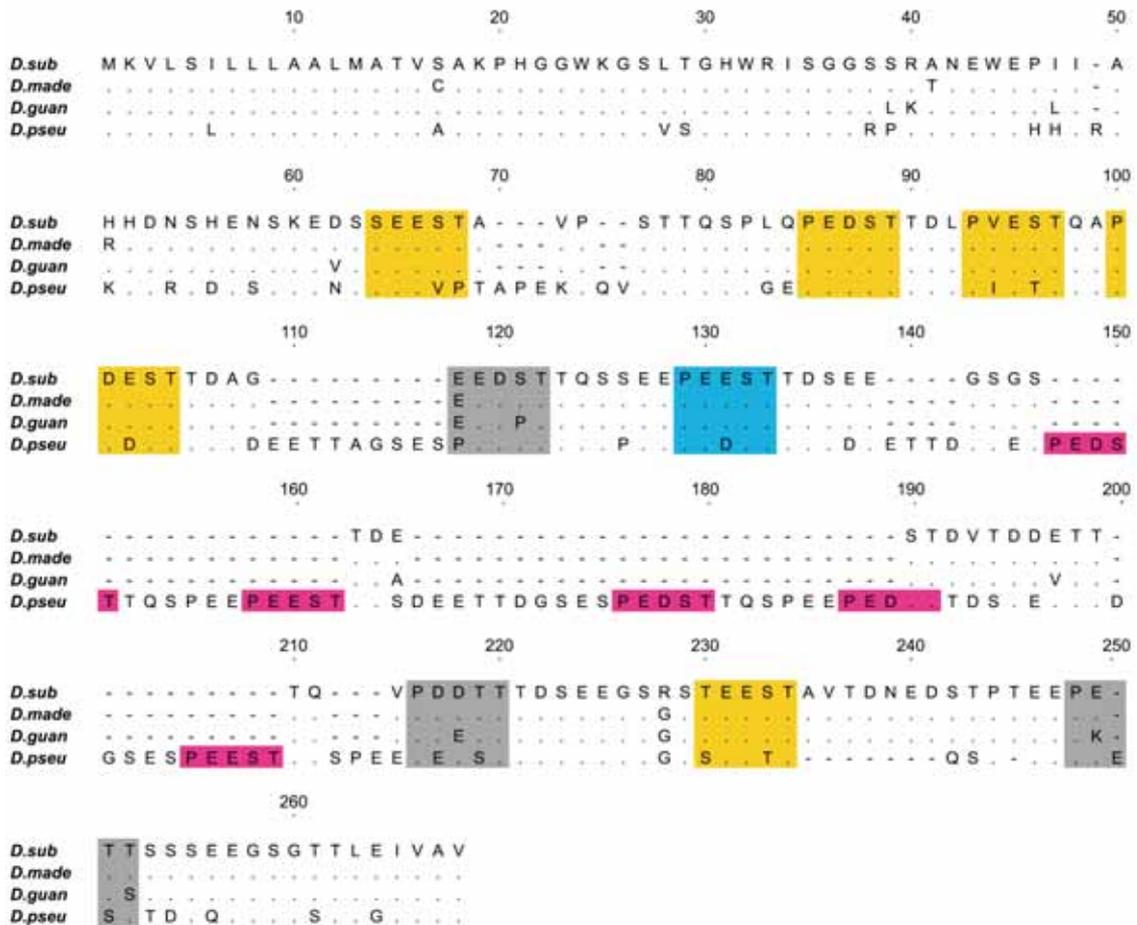


Figura I.14. Alineamiento múltiple de la secuencia aminoacídica de la proteína codificada por el gen *Fst* en el grupo *obscura* de *Drosophila*. Los puntos (.) indican aminoácidos idénticos a los de la secuencia de referencia y los guiones (-) deleciones de aminoácidos respecto a esta secuencia. En amarillo se indican los motivos PEEST que presentan un cambio en uno de los residuos en la secuencia de referencia. En fucsia, las inserciones de motivos con respecto a esta secuencia. En azul, los motivos PEEST en *D.sub* que presentan un cambio en uno de los residuos en alguna de las especies restantes. En gris, los motivos PEEST perfectos o con un cambio en uno de los residuos en *D.pseu* y que presentan más de un cambio en el resto de las especies. *D.sub* *D. subobscura*; *D.made*, *D. madeirensis*; *D.guan*, *D. guanche*; *D.pseu*, *D. pseudoobscura*.

I.2.2 DISCUSIÓN

I.2.2.1 DIVERGENCIA NUCLEOTÍDICA Y AMINOACÍDICA DEL GEN *Fst* EN *D. MELANOGASTER* Y *D. SUBOBSCURA*

La divergencia en el gen *Fst* se ha analizado independientemente en los grupos *melanogaster* y *obscura* de *Drosophila*. El motivo de no haber realizado un análisis conjunto es la dificultad de alinear con una cierta seguridad las secuencias del gen entre ambos grupos de especies debido principalmente al elevado número de adiciones/delecciones que se han producido en la región de los motivos PEEST durante la divergencia. En la figura I.15 se presenta un posible alineamiento de la proteína codificada por el gen *Fst* entre *D. melanogaster* y *D. subobscura*. La similitud más elevada de la secuencia aminoacídica de ambas especies se encuentra en los primeros 26 aminoácidos de la región N-terminal de la proteína. El alineamiento del resto de la proteína es muy dudoso y puede considerarse tan solo una aproximación. Sin embargo, este alineamiento entre *D. melanogaster* y *D. subobscura* se utilizó para estimar la divergencia sinónima y no sinónima en la región codificadora del gen *Fst* entre estas especies. La estima de divergencia sinónima ($K_s = 1.232$) es comparable a las estimas encontradas en 17 genes secuenciados en *D. melanogaster* y *D. subobscura* recopilados en Munté et al (1997) y que oscilan entre 0.536 en el gen *Sxl* y 1.586 en el gen *Adhr*. Por el contrario, el gen *Fst* presenta la estima de divergencia no sinónima ($K_a = 0.512$) más elevada en comparación a la detectada en los otros genes secuenciados en ambas especies. Estas estimas oscilan entre 0.021 en el gen *Antp* y 0.405 en el gen *Cp15*. Este resultado indicaría que el gen *Fst* está evolucionando rápidamente en las posiciones no sinónimas.

La evolución rápida del gen *Fst* se manifiesta tanto por la elevada divergencia no sinónima como por el acumulo de inserciones/delecciones. De hecho, una de las características particulares del gen *Fst* es la presencia de una serie de repeticiones (PEEST) que se ubican en la región C-terminal de la proteína. El número de repeticiones en *D. melanogaster* es de 10 pero en *D. subobscura* se reduce a 1. De igual forma, el número de motivos que difieren en uno de los residuos (PEDST, TEEST, PVEST, etc) en *D. melanogaster* (13) es mayor que en *D. subobscura* (9). Esta variación en el número de motivos no sólo se detecta entre los grupos *melanogaster* y *obscura* sino también dentro de cada uno de los grupos. La longitud de la proteína en las

El número de prolina varía tanto entre especies como dentro de especie ya que se ha detectado polimorfismo en esta característica en *D. melanogaster*. La forma mayoritaria en esta especie presenta 10 prolina consecutivas, mientras que este número es 7 en *D. simulans* y *D. yakuba*, 5 en *D. mauritiana*, y 9 en *D. erecta*. Sorprendentemente, las especies del grupo *obscura* han perdido esta característica. Se requeriría un análisis funcional comparativo entre las especies de los grupos *melanogaster* y *obscura* para determinar si la pérdida de los residuos de prolina afecta a la función de la proteína FROST.

La divergencia del gen *Fst* en el grupo *melanogaster* se analizó en un total de 5 especies. La estima de la divergencia sinónima entre *D. melanogaster* y *D. simulans* ($K_s = 0.144$) puede considerarse relativamente elevada si se compara con la estima promedio de la divergencia nucleotídica silenciosa de 19 genes autosómicos analizados en ambas especies ($K_{sil} = 0.108$) (Begun y Whitley 2000). De hecho, únicamente 3 de los 19 genes están evolucionando más rápidamente en las posiciones sinónimas que el gen *Fst*. Un argumento similar es válido al considerar la divergencia no sinónima. La estima de K_a en el gen *Fst* de 0.035 es superior a la estima promedio para los mismos genes ($K_a = 0.011$) (Begun 2002). También en este caso sólo 3 de los 19 genes presentan estimas de K_a superiores a las del gen *Fst*.

La relación K_a/K_s es siempre inferior en las comparaciones entre las especies del grupo *melanogaster* que en las comparaciones entre especies del grupo *obscura*. De hecho, la estima promedio de la relación K_a/K_s en las especies del grupo *melanogaster* y del grupo *obscura* es 0.198 y 0.325, respectivamente. Este resultado indicaría que las constricciones funcionales de la proteína FROST son menores en el grupo *obscura*, lo que comporta una menor selección purificadora en este grupo de especies.

En general, los árboles filogenéticos reconstruidos de acuerdo a la divergencia del gen *Fst* están de acuerdo con las relaciones filogenéticas de las especies analizadas. Tan solo se detecta una discrepancia en el árbol reconstruido en base a la divergencia sinónima en el grupo *melanogaster* (figura I.11B) en el que *D. simulans* aparece más estrechamente relacionada con *D. melanogaster* que con *D. mauritiana*. Sin embargo, está generalmente aceptado que *D. simulans* es evolutivamente más cercana a *D. mauritiana*. Por otro lado, los árboles filogenéticos de la figura I.11 indican una estrecha relación entre *D. yakuba* y *D. erecta*. Aunque *D. erecta* se consideraba la especie *outgroup* de todas las analizadas, estudios recientes (Ko et al. 2003) ya apuntan hacia una estrecha relación entre *D. yakuba* y *D. erecta*, tal como se observa de acuerdo

a la divergencia del gen *Fst*. Los árboles filogenéticos de las figuras I.11A y I.11C muestran que la rama que conduce a *D. erecta* es considerablemente más larga que la que conduce a *D. yakuba*. La filogenia obtenida permite la aplicación del *relative rate test* utilizando cualquier especie del complejo *melanogaster* como *outgroup*. Utilizando *D. melanogaster* como especie *outgroup* se detectó una diferencia altamente significativa en la tasa de sustitución nucleotídica entre los linajes de *D. yakuba* y de *D. erecta* causada por una aceleración en la tasa de sustitución no sinónima en el linaje de *D. erecta*. De las 32 sustituciones aminoacídicas fijadas en el linaje de *D. erecta*, 12 están ubicadas en la región N-terminal de la proteína entre las posiciones 25 y 75. La mayoría de estos 12 reemplazamientos, un total de 8, son radicales con valores elevados de distancia físico-química entre los aminoácidos. Por lo tanto, la selección positiva podría haber causado la fijación de algunos de estos reemplazamientos posiblemente debido a cambios beneficiosos en la especificidad de la proteína.

Las estimas de divergencia nucleotídica en las regiones no codificadoras de todas las especies son bastante inferiores a las detectadas en las posiciones sinónimas de la región codificadora. Esta restricción en las posiciones no codificadoras es más evidente en la región 5' flanqueante del gen *Fst* en las especies del grupo *melanogaster* y en la región 3' en las especies del grupo *obscura*. La diferente longitud de las regiones 5' y 3' analizadas en ambos grupos de especies podría explicar esta discrepancia. No obstante, la elevada diferencia entre la divergencia sinónima y la divergencia de las regiones no codificadoras indica que la selección purificadora está manteniendo regiones importantes posiblemente involucradas en la regulación del gen *Fst*. Sin embargo, hasta el momento las regiones promotoras y reguladoras de este gen no han sido caracterizadas.

Las estimas del sesgo en el uso de codones sinónimos son más elevadas en las especies del grupo *melanogaster* que en las del grupo *obscura*. Se ha propuesto que el ambiente recombinacional puede afectar al uso de codones sinónimos (Kliman y Hey 1993; Munté et al. 1997) en el sentido de que el *codon bias* es inferior en las regiones de baja recombinación debido a una menor eficacia de la selección débil en el mantenimiento de los codones preferentes. Sin embargo, el gen *Fst* se encuentra en una región con niveles reducidos de recombinación en, por ejemplo, *D. melanogaster* en comparación con *D. subobscura*. Por lo tanto, las diferencias en el ambiente recombinacional no parecen poder explicar las diferencias detectadas en el uso de codones sinónimos entre las especies de los grupos *melanogaster* y *obscura*. Aunque el

menor *codon bias* en las especies del grupo *obscura* podría indicar una mayor divergencia sinónima en estas especies, este hecho no se refleja en los datos obtenidos ya que, como se ha indicado, la relación K_a/K_s es superior en las especies del grupo *obscura* que en las del grupo *melanogaster*. Sin embargo, los menores niveles de las constricciones funcionales en algunas proteínas permiten acumular un mayor número de sustituciones no sinónimas causando un incremento en el valor de la relación K_a/K_s .

CAPÍTULO II

EL GEN *DROSOPHILA COLD ACCLIMATION* (*Dca*)

II.1. POLIMORFISMO NUCLEOTÍDICO DEL GEN *Dca*

II.1.1 RESULTADOS

II.1.1.1 POLIMORFISMO NUCLEOTÍDICO DEL GEN *Dca* EN *D. MELANOGASTER*

La variabilidad nucleotídica del gen *Dca* fue analizada en 21 líneas de *D. melanogaster* capturadas en dos poblaciones naturales: Montemayor y Sant Sadurní d'Anoia. La región secuenciada comprende aproximadamente 1.6 kb e incluye la totalidad de la región codificadora (aproximadamente 1 kb) y parte de las regiones 5' y 3' flanqueantes con una longitud de alrededor de 0.5 kb y 50 pb, respectivamente.

Diferenciación genética entre poblaciones

La diferenciación genética entre las poblaciones de Montemayor y Sant Sadurní d'Anoia fue estimada usando el estadístico F_{st} (Hudson et al. 1992). El valor observado al comparar las dos poblaciones analizadas ($F_{st} = 0.126$) no difiere significativamente del esperado bajo el modelo neutro en muestras aleatorias de una población panmítica sin subdivisión (figura II.1). Por lo tanto, ambas poblaciones se consideraron conjuntamente en los análisis posteriores.

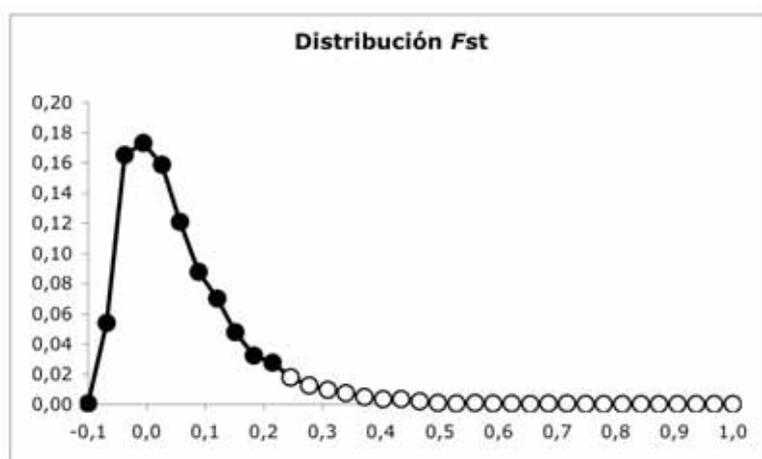


Figura II.1. Distribución del estadístico F_{st} en una población panmítica de 21 individuos (ver material y métodos).

Análisis del polimorfismo nucleotídico

El alineamiento múltiple de las 21 líneas de *D. melanogaster* presenta un total de 1577 posiciones que se reducen a 1556 al eliminar las posiciones con *gaps*. En el alineamiento se detectaron un total de 7 polimorfismos por inserción/delección y 42 polimorfismos por cambio nucleotídico (figura II.2). Seis de los 7 polimorfismos por inserción/delección se detectaron en la región 5' flanqueante donde también se observaron 14 polimorfismos por cambio nucleotídico, uno de los cuales presenta una variante única. La región codificadora presenta 23 polimorfismos por cambio nucleotídico, de los cuales 17 son sinónimos y 6 no sinónimos. Cinco de las 17 posiciones polimórficas presentan variantes únicas. En la región codificadora no se detectaron polimorfismos por inserción/delección. En el intrón se detectaron 2 polimorfismos por cambio nucleotídico y uno por inserción/delección. En la región 3' flanqueante se identificaron 3 polimorfismos por cambio nucleotídico, uno de los cuales presenta una variante única.

Análisis del polimorfismo aminoacídico

El alineamiento múltiple de la secuencia aminoacídica de la proteína codificada por el gen *Dca* en las 21 líneas estudiadas de *D. melanogaster* se muestra en la figura II.3. El alineamiento completo contiene un total de 303 aminoácidos. Se observaron 6 polimorfismos no sinónimos de los cuales 3 presentan variantes únicas. Entre los polimorfismos aminoacídicos detectados, se encontraron tanto cambios de aminoácidos conservativos como radicales. Por ejemplo, los cambios lisina/arginina (Lys/Arg), valina/isoleucina (Val/Ile) y ácido aspártico/glutamina (Asp/Glu) en las posiciones 713, 1216 y 1227, respectivamente, pueden clasificarse como conservativos debido a que no producen cambio ni en la carga ni en la polaridad del aminoácido. Por el contrario, los cambios alanina/ácido aspártico (Ala/Asp) y valina/ácido aspártico (Val/Asp) en las posiciones 947 y 1519, respectivamente, son radicales puesto que alteran las propiedades físico-químicas del aminoácido.

		713	947	1057	1216	1227	1519
CN1	P1	Lys	Ala	Asp	Val	Asp	Val
M13	
CN23	
CN26	
CN34	P2	Arg
M2		Arg
CN9	P3	Arg	Asp
M47		Arg	Asp
M54		Arg	Asp
CN22		Arg	Asp
CN24		Arg	Asp
M40	P4	.	Asp	.	.	Glu	.
CN7	P5	.	.	Tyr	.	Glu	.
M11	P6	.	Asp
M28		.	Asp
M55		.	Asp
M59		.	Asp
M36		.	Asp
CN41	P7	.	Asp	.	.	Glu	.
M66	P8	.	Asp	.	Ile	.	Asp
CN18	P9	Arg	Asp	.	.	Glu	.
AB029491		Arg	Asp
cDNA		.	Asp
<i>D. simulans</i>		.	Asp	.	.	Glu	.

Figura II.3. Polimorfismo aminoacídico de la proteína codificada por el gen *Dca* en las 21 de *D. melanogaster* líneas secuenciadas de las poblaciones de Montemayor (M) y Sant Sadurn d'Anoia (CN). P# variante proteica. Ver el pie de la figura II.2 para detalles.

Diversidad haplotípica

Las 21 líneas de *D. melanogaster* analizadas se agrupan en 14 haplotipos al considerar todos los polimorfismos por cambio nucleotídico detectados en la región estudiada. El número de haplotipos identificados no se ve afectado si se consideran asimismo los polimorfismos por inserción/delección. De los 14 haplotipos identificados, once corresponden a líneas únicas, dos de ellos están presentes en 4 líneas y el restante es común a 2 líneas. La diversidad haplotípica es $Hd = 0.938$.

La región 5' flanqueante muestra un estructura haplotípica muy acusada ya que en ella sólo se detectan 9 haplotipos en las 21 líneas analizadas. De hecho, en esta región la diversidad haplotípica es 0.781 (*vs* 0.938 en toda la región estudiada) con un haplotipo compartido por 9 líneas.

Cuando se consideran únicamente los polimorfismos observados en la región codificadora, las 21 líneas analizadas se agrupan en 14 haplotipos que codifican para 9 variantes proteicas (figura II.3). En esta figura se observa que las variantes P3 y P6 son las que presentan una mayor frecuencia (un 23.8% de las líneas analizadas). Una de las variantes (P1) alcanza una frecuencia del 19% mientras que las restantes (P2, P4, P5, P7, P8 y P9) aparecen con una frecuencia inferior al 10%. Entre estas últimas, la variante P2 se encontró en 2 líneas, mientras que las restantes están presentes en una sola línea.

Desequilibrio de ligamiento y recombinación

El grado de desequilibrio de ligamiento se analizó para las posiciones polimórficas informativas. Los 35 polimorfismos informativos detectados permitieron hacer un total de 595 comparaciones, de las que 174 (29.2%) fueron significativas con el test χ^2 y 25 (4.2%) permanecieron significativas tras aplicar la corrección de Bonferroni para múltiples tests. A pesar del relativamente elevado porcentaje de comparaciones significativas, el nivel global de desequilibrio en el total de la región analizada no fue significativo de acuerdo a los estadísticos Z_s ($Z_s = 0.179$; $P = 0.668$), Z_A ($Z_A = 0.309$; $P = 0.27$), Wall's B (Wall's $B = 0.117$; $P = 0.64$) y Wall's Q (Wall's $Q = 0.228$; $P = 0.463$) bajo la hipótesis conservativa de no recombinación. El número mínimo de eventos de recombinación (R_m) detectados fue de 9: tres en la región 5' flanqueante, 5 en la región codificadora y 1 entre estas dos regiones.

Se detectó un acumulo de desequilibrios de ligamiento significativos en la región 5' flanqueante. Así, 55 de las 78 comparaciones (70%) realizadas entre los polimorfismos de esta región fueron significativas con el test χ^2 y 31 de ellas (39.7%) permanecieron significativas tras aplicar la corrección de Bonferroni. Este resultado está de acuerdo con la fuerte estructura haplotípica presente en esta región.

Por el contrario, en la región codificadora (incluyendo el intrón) 43 de las 190 (22.6%) comparaciones realizadas son significativas por el test de χ^2 y tan solo 9 de ellas (4.7%) permanecen significativas tras aplicar la corrección de Bonferroni.

Diversidad nucleotídica

Los niveles del polimorfismo nucleotídico fueron estimados para distintos tipos de posiciones (tabla II.1). La diversidad nucleotídica (π) para toda la región analizada fue 0.009. Esta estima es ligeramente inferior a la de 0.015 obtenida para las posiciones silenciosas (regiones 5' y 3' flanqueantes, intrón y posiciones sinónimas de la región codificadora). Dentro de la región codificadora, la estima para las posiciones no sinónimas es 0.003, mientras que para las posiciones sinónimas es 0.026. En la región no codificadora (regiones 5' y 3' flanqueantes e intrón) la diversidad nucleotídica es 0.011, estima bastante inferior a la diversidad nucleotídica sinónima ($\pi_s = 0.026$). Las estimas de θ fueron en general inferiores a las estimas de π , resultado que indica un exceso de polimorfismos con variantes segregando a frecuencias intermedias en la región *Dca*.

Tabla II.1. Estimaciones del polimorfismo nucleotídico del gen *Dca* en *D. melanogaster*

Región	Nº de Sitios	S	π	θ
No Codificadora	647	19	0.011	0.008
5'	541	14	0.010	0.007
Intrón	58	2	0.014	0.009
3'	48	3	0.014	0.017
Codificadora	909	23	0.008	0.007
Silenciosa	862.466	36	0.015	0.012
Sinónima	215.468	17	0.026	0.022
No sinónima	693.532	6	0.003	0.002
Total	1556	42	0.009	0.008

Test del neutralismo

Los test de neutralismo propuestos por Tajima (1989) y Fu y Li (1993) se aplicaron para determinar si el patrón del polimorfismo observado en la región *Dca* se ajusta a las predicciones del neutralismo (tabla II.2). Estos tests se realizaron independientemente para diferentes tipos de posiciones. Los estadísticos propuestos para aplicar los distintos tests fueron positivos para todos los tipos de posiciones excepto para las no sinónimas. Los resultados de aplicar los distintos tests fueron siempre no significativos a excepción de los test de Fu and Li en la región 5' flanqueante ($P < 0.05$), región en la que el test de Tajima fue marginalmente significativo. Estos resultados indican que el patrón del polimorfismo observado en la región analizada se ajusta a las predicciones del neutralismo, a excepción de la región 5' flanqueante donde existe una desviación hacia un exceso de variantes segregando a frecuencias intermedias.

Tabla II.2. Test del neutralismo del gen *Dca* en *D. melanogaster*

Test	No codificadora	5'	Silenciosa	No sinónima	Total
Tajima D	1.270	1.561	1.124	- 0.181	0.939
<i>P(D < D_{obs})</i>	0.928	0.054 [#]	0.896	0.510	0.859
Fu y li D*	1.007	1.135	1.048	- 0.887	0.734
<i>P(D* < D*_{obs})</i>	0.951	0.038*	0.894	0.305	0.837
Fu y Li F*	1.263	1.465	1.249	- 0.794	0.929
<i>P(F* < F*_{obs})</i>	0.936	0.043*	0.935	0.244	0.852

Probabilidad de obtener un valor del test estadístico menor que el observado tras simulación de coalescencia.

* $P < 0.05$

[#] $0.05 < P < 0.1$

El test de MacDonald y Kreitman (1991) se aplicó para determinar si la relación entre el número de polimorfismos dentro de especie y el número de diferencias fijadas entre especies es equivalente para los cambios sinónimos y no sinónimos. Para realizar la comparación interespecífica se utilizó *D. simulans*. No se detectaron diferencias significativas entre ambas relaciones (tabla II.3).

Tabla II.3. Test de MK (1991) en el gen *Dca* de *D. melanogaster*

	<u>Polimorfismo en <i>D. melanogaster</i></u>	<u>Diferencias Fijadas con <i>D. simulans</i></u>
No sinónimo	6	11
Sinónimo	17	23
		G value = 0.260 P = 0.610
		NI = 0.738

Genealogía de los alelos

La figura II.4 muestra la genealogía de los diferentes alelos de las poblaciones de Montemayor y Sant Sadurní d'Anoia incluyendo también la línea del proyecto de secuenciación del genoma de *D. melanogaster* así como la secuencia publicada del cDNA (Goto 2000). Para este análisis se utilizó la secuencia de *D. simulans* como especie *outgroup*. En el árbol se pueden identificar los dos casos de haplotipos compartidos por cuatro líneas (CN1-M13-CN23-CN26 y M11-M26-M55-M59) y las dos líneas que comparten otro haplotipo (CN22-CN24). No se observan agrupaciones de alelos por localidad, lo que apoya el hecho de que las poblaciones analizadas no se encuentran genéticamente diferenciadas.

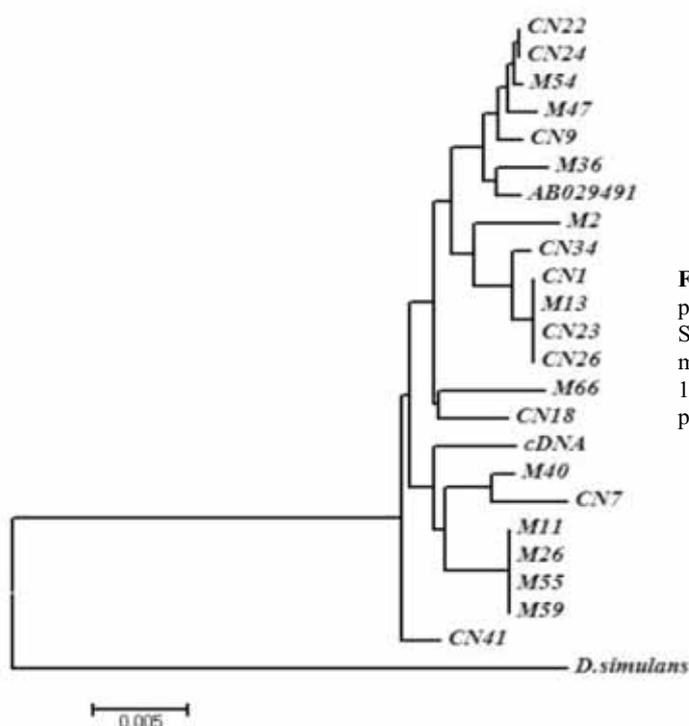


Figura II.4. Genealogía de los alelos de las poblaciones de Montemayor (M) y Sant Sadurní d'Anoia (CN) reconstruida por el método de *neighbor-joining* (Saitou y Nei, 1987) usando la distancia de Kimura de 2 parámetros (Kimura 1980).

II.1.1.2 POLIMORFISMO NUCLEOTÍDICO DEL GEN *DCA* EN *D. SUBOBSCURA*

Se analizó un fragmento de aproximadamente 1.6 Kb en 13 líneas de *D. subobscura* de una población natural de Galicia (España). El fragmento analizado contiene toda la región codificadora (968 pb) del gen *Dca* y parte de las regiones 5' y 3' flanqueantes con una longitud aproximada de 0.3 y 0.5 kb, respectivamente.

Ubicación citológica del gen *Dca*

Los resultados de la hibridación *in situ* del gen *Dca* en los cromosomas politénicos de *D. subobscura* muestra una señal de hibridación ubicada en la banda citológica 86A del cromosoma O cerca al punto de rotura distal de la inversión 7 (77B/C-85E) (figura II.5).



Figura II.5. Hibridación *in situ* del gen *Dca* en cromosomas politénicos de *D. subobscura*. La flecha indica la señal de hibridación.

Análisis del polimorfismo nucleotídico

El alineamiento múltiple de las 13 líneas de *D. subobscura* comprende un total de 1808 posiciones. Sin embargo, el número de posiciones se reduce a 1660 después de eliminar aquellas que presentan *gaps* en una o varias de las líneas. Se detectaron tanto polimorfismos por inserción/delección como polimorfismos por cambio nucleotídico. El número total de estos polimorfismos en toda la región secuenciada fue de 14

polimorfismos por inserción/delección y 84 polimorfismos por cambio nucleotídico (figura II.6). En dos posiciones polimórficas (una en la región codificadora y una en la región 3' flanqueante) se detectaron segregando 3 variantes, por lo que el número mínimo de mutaciones es 86.

En la región 5' flanqueante se observaron 4 de los 14 polimorfismos por inserción/delección y 21 posiciones polimórficas por cambio nucleotídico, de las cuales 17 presentan variantes únicas. La región codificadora presenta 33 polimorfismos por cambio nucleotídico y no se detectó ningún polimorfismo por inserción/delección. De los 33 polimorfismos detectados, 11 son no sinónimos y 22 sinónimos. Un total de 11 polimorfismos de estos 33 presentan variantes únicas. En el intrón se observaron 9 polimorfismos por cambio nucleotídico, de los cuales 4 presentan variantes únicas y un polimorfismo por inserción/delección. En la región 3' flanqueante se detectaron un total de 22 polimorfismos por cambio nucleotídico, de los que 15 presentan variantes únicas. Además, en esta región se identificaron nueve polimorfismos por inserción/delección.

Análisis del polimorfismo aminoacídico

En la figura II.7 se muestra el alineamiento múltiple de la secuencia aminoacídica de la proteína codificada por el gen *Dca* en las 13 líneas de *D. subobscura*. Este alineamiento incluye un total 303 aminoácidos. Se detectaron un total de 11 polimorfismos no sinónimos que representan 10 polimorfismos aminoacídicos, ya que dos de los cambios no sinónimos (posiciones 731 y 732) afectan al mismo codón. Cinco de los 11 polimorfismos no sinónimos presentan variantes únicas.

De los polimorfismos aminoacídicos detectados se encontraron tanto cambios conservativos como radicales. Por ejemplo, los cambios fenilalanina/triptófano (Phe/Trp) y valina/isoleucina (Val/Ile) en las posiciones 731 y 976, respectivamente, pueden considerarse como conservativos puesto que no varían ninguna de las propiedades físico-químicas del aminoácido. Por el contrario, algunos cambios como glutamina/valina (Glu/Val) y lisina/metionina (Lys/Met) en las posiciones 392 y 1244, respectivamente, alteran la polaridad del aminoácido y, por lo tanto, pueden considerarse como radicales. Al tener en cuenta las secuencias de *D. madeirensis* y *D. guanche*, todos los aminoácidos presentes en una única línea son derivados.

		392	731-732	749	860	959	976	1039	1118	1244	1282
LS89.754	P1	Glu	Phe	Asp	Leu	Phe	Val	Pro	Leu	Lys	Pro
LS89.669	P2	Val	Trp	Gly	His	Tyr	Ile
LS89.567	P3	Val	Trp	.	His	Tyr	Ile
S89.4		Val	Trp	.	His	Tyr	Ile
LS89.517	P4	Val	Trp	Gly	His	Tyr	Ile	Thr	.	.	.
AU89.918	P5	Val	Trp	Gly	His	Tyr	Ile	.	Gln	.	.
AU89.898	P6	Val	Trp	.	His	Tyr	Ile	.	.	Met	.
S89.174	P7	Val	Trp	.	.	.	Ile	.	.	.	Ser
ES89.484	P8	Val	Trp
AU89.1002	P9	Val
AU89.986		Val
AU89.784		Val
AU89.817	P10	Val	.	.	His	Tyr	Ile
<i>D. guanche</i>		Val	.	.	His	Tyr	Ile
<i>D. madeirensis</i>		Val	.	.	His	Tyr

Figura II.7. Polimorfismo aminoacídico de la proteína codificada por el gen *Dca* en las 13 líneas de *D. subobscura* secuenciadas de la población de Galicia. P#, variante proteica. Ver pie de la figura II.6 para detalles.

Diversidad haplotípica

Las 13 líneas de *D. subobscura* analizadas se agrupan en 13 haplotipos al considerar todos los polimorfismos por cambio nucleotídico detectados en la región estudiada. Es decir, cada línea corresponde a un haplotipo y por lo tanto la diversidad haplotípica es $Hd = 1$.

Cuando se consideran únicamente los polimorfismos detectados en la región codificadora, las 13 líneas de *D. subobscura* también corresponden a 13 haplotipos que representan 10 variantes proteicas. En la figura II.7 se observa que las variantes proteicas P3 y P9 presentan la mayor frecuencia (están presentes en el 15,4% y 23,1% de las líneas analizadas, respectivamente). El resto de variantes (P2, P4-P8 y P10) se encuentran en líneas individuales (con una frecuencia menor del 10%).

Desequilibrio de ligamiento y recombinación

El grado de desequilibrio de ligamiento se analizó entre las 61 posiciones polimórficas informativas detectadas. Un total de 75 de las 595 comparaciones realizadas (12.6%) fueron significativas por el test χ^2 , pero ninguna de ellas permaneció significativa tras aplicar la corrección de Bonferroni. Se detectaron 13 eventos de recombinación en toda la región analizada: 1 en la región 5' flanqueante, 7 en la región codificadora, 1 entre la región codificadora y el intrón, 1 entre la región codificadora y la región 3' flanqueante y 3 en la región 3' flanqueante.

Diversidad nucleotídica

Los niveles de la variabilidad nucleotídica intraespecífica fueron estimados para distintos tipos de posiciones (tabla II.4). La diversidad nucleotídica silenciosa ($\pi_{\text{sil}} = 0.02$) es similar a la detectada en *D. melanogaster* ($\pi_{\text{sil}} = 0.015$, tabla II.1). Dentro de la región codificadora, la estima para las posiciones sinónimas (π_{s}) es 0.033, mientras que para las posiciones no sinónimas (π_{a}) es de 0.006. La estima detectada en la región no codificadora (regiones 5' y 3' flanqueantes e intrón) de 0.016 es de aproximadamente la mitad de la diversidad nucleotídica para las posiciones sinónimas ($\pi_{\text{s}} = 0.033$).

Tabla II.4. Estimaciones del polimorfismo nucleotídico del gen *Dca* en *D. subobscura*

Región	Nº de Sitios	S (η)	π	θ
No Codificadora	751	52(53)	0.016	0.023
5'	308	21	0.013	0.022
Intron	59	9	0.054	0.050
3'	384	22(23)	0.012	0.020
Codificadora	909	33	0.012	0.012
Silenciosa	975.859	74	0.020	0.025
Sinónima	224.859	22	0.033	0.031
No sinónima	684.141	11	0.006	0.005
Total	1660	83(85)	0.014	0.017

η , número de mutaciones

Test del neutralismo

Los tests de neutralismo propuestos por Tajima (1989) y Fu y Li (1993) se aplicaron para determinar si el patrón de polimorfismo detectado en la región *Dca* en *D. subobscura* se ajusta al esperado bajo el neutralismo (tabla II.5). Estos tests fueron realizados independientemente para diferentes tipos de posiciones. Los estadísticos propuestos para aplicar estos tests son negativos en todos los casos a excepción del estadístico D de Tajima para las posiciones no sinónimas de la región codificadora. Sin embargo, solamente los tests realizados en la región 5' flanqueante son significativos ($P < 0.05$), mientras que son marginalmente significativos ($0.05 < P < 0.1$) los tests realizados en la región no codificadora. Estos resultados sugieren que el patrón del polimorfismo observado en la región analizada concuerda con el esperado bajo el neutralismo, excepto en la región 5' flanqueante que presenta un exceso de variantes segregando a baja frecuencia debido a la abundancia de variantes únicas.

Tabla II.5. Tests del neutralismo en el gen *Dca* de *D. subobscura*

Test	No codificadora	5'	Silenciosa	No sinónima	Total
Tajima D	- 1.311	- 1.690	- 0.865	0.256	- 0.716
$P(D < D_{obs})$	0.080	0.037*	0.183	0.658	0.248
Fu y Li D*	- 1.711	- 2.056	- 1.211	- 0.437	- 1.123
$P(D^* < D_{obs}^*)$	0.070	0.045*	0.141	0.384	0.136
Fu y Li F*	- 1.835	- 2.238	- 1.280	- 0.290	- 1.161
$P(F^* < F_{obs}^*)$	0.070	0.025*	0.132	0.387	0.177

Probabilidad de obtener un valor del test estadístico menor que el observado tras simulación de coalescencia sin recombinación.

* $P < 0.05$.

El test de MacDonald y Kreitman (1991) se aplicó para determinar si la relación entre el número de polimorfismos dentro de especie y el número de diferencias fijadas entre especies es equivalente para los cambios sinónimos y no sinónimos. Para la comparación interespecífica se utilizaron *D. guanche* o *D. madeirensis* (tabla II.6). No se detectaron diferencias significativas entre ambas relaciones en ninguna de las dos comparaciones.

Tabla II.6. Test de MK (1991) en el gen *Dca* de *D. subobscura*

	Diferencias Fijadas		
	Polimorfismo en <i>D. subobscura</i>	<i>D. guanche</i>	<i>D. madeirensis</i>
No sinónimo	11	5	0
Sinónimo	22	19	7
		G value = 1.099	Fisher's exact test
		<i>P</i> = 0.294	<i>P</i> = 0.159
		NI = 1.9	

El test de HKA (Hudson et al. 1987) se aplicó para determinar si el nivel del polimorfismo silencioso detectado en la región *Dca* está correlacionado con el nivel de divergencia silenciosa observado entre *D. subobscura* y *D. guanche*. Para aplicar este test, los niveles de polimorfismo y divergencia silenciosa de la región *Dca* se compararon con los niveles detectados en la región *rp49* de una muestra de Galicia con la ordenación O_{3+4} (Rozas et al. 1995) asumiendo que esta región evoluciona siguiendo el modelo neutro. El valor obtenido del estadístico χ^2 en el test HKA no es significativo ($\chi^2 = 0.086$, *P* = 0.768).

Genealogía de los alelos

La genealogía de los diferentes alelos de la población de Galicia de *D. subobscura* se muestra en la figura II.8. La reconstrucción del árbol genealógico se realizó a partir de la variación detectada en toda la región analizada. Para este análisis se utilizó la secuencia *D. madeirensis* como *outgroup*.

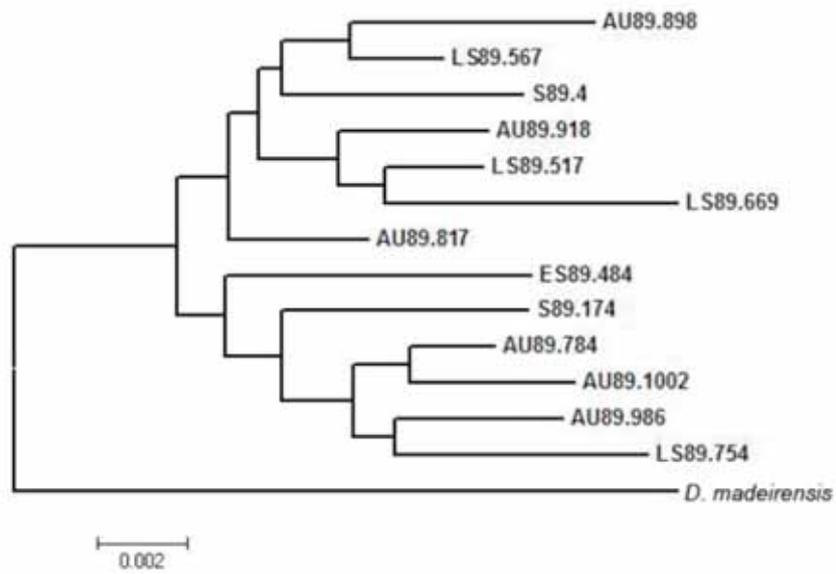


Figura II.8. Genealogía de los 13 alelos analizados de *D. subobscura* reconstruida por el método *neighbor-joining* (Saitou y Nei 1987) utilizando la distancia de Kimura de 2 parámetros (Kimura 1980).

II.1.2 DISCUSIÓN

II.1.2.1 DIVERSIDAD NUCLEOTÍDICA DEL GEN *Dca* EN *D. MELANOGASTER* Y *D. SUBOBSCURA*

En los últimos años se han realizado varios estudios con el propósito de detectar genes que se expresen durante la exposición y recuperación a las bajas temperaturas en *Drosophila*, especialmente en *D. melanogaster*. El gen *Dca* ha sido identificado como un gen que se sobreexpresa durante la aclimatación a 15° C en *D. melanogaster* (Goto 2000). Posteriormente, se han realizado análisis con *microarrays* (Qin et al. 2005) y RT-PCR cuantitativa (Sinclair et al. 2007) para determinar los posibles cambios en la abundancia de los transcritos en individuos de *D. melanogaster* después de la exposición a 0°C y a otros tipos de estrés (deshidratación e inanición). Estos estudios han detectado una reducción en la abundancia del mRNA del gen *Dca* en la respuesta al frío. Estas discrepancias en los niveles de expresión de este gen han sido explicadas de acuerdo a la diferencia en la temperatura empleada en los diferentes estudios (15°C vs 0°C). La posible función de la proteína DCA se ha relacionado con la regulación del Ca²⁺ debido a su similitud con la proteína SMP-30 de los mamíferos (Goto 2000). Por lo tanto, se ha sugerido que los cambios necesarios en la expresión del gen *Dca* para la regulación del Ca²⁺ durante la aclimatación a bajas temperaturas y durante la recuperación después de un choque frío son diferentes.

En el presente estudio el polimorfismo nucleotídico en la región *Dca* de *D. melanogaster* fue analizado conjuntamente para las poblaciones naturales de Montemayor y Sant Sadurní d'Anoia debido a la falta de diferenciación genética entre ellas. Al igual que en el gen *Fst*, las diferencias climáticas que presentan las poblaciones analizadas no parecen ser lo suficientemente importantes para determinar una diferenciación genética en el gen *Dca*.

A diferencia de lo detectado en el gen *Fst*, los niveles de diversidad nucleotídica silenciosa del gen *Dca* en *D. melanogaster* ($\pi_{\text{sil}} = 0.015$) y *D. subobscura* ($\pi_{\text{sil}} = 0.02$) son similares, aunque ligeramente superiores en esta última especie. Este resultado refuerza que el bajo nivel de polimorfismo del gen *Fst* en *D. melanogaster* se debe a su localización en

una zona de baja recombinación. Por otro lado, el gen *Dca* está asociado pero no afectado por inversiones cromosómicas que segregan como polimórficas en ambas especies. En *D. melanogaster* el gen se encuentra ubicado en la banda 88D2 del brazo cromosómico 3R, a aproximadamente una sección cromosómica del punto de rotura proximal de la inversión In(3R)P (89C2-96A1). Asimismo, en *D. subobscura* se ubica en la banda 86A relativamente cercano al punto de rotura distal de la inversión 7 (77B/C-85E). No obstante, esta asociación del gen *Dca* con inversiones cromosómicas no parece afectar al nivel de variación del gen en ninguna de las dos especies.

El valor del nivel de polimorfismo silencioso del gen *Dca* en *D. subobscura* ($\pi_{\text{sil}} = 0.02$) se encuentra entre los más elevados detectados en la especie con un valor promedio para los loci autosómicos de $\pi_{\text{sil}} = 0.0122$ (tabla I.8). Tan solo los genes *AcpH-1* ($\pi_{\text{sil}} = 0.0210$ en la ordenación O_{3+4}) y *Xdh* ($\pi_{\text{sil}} = 0.0337$) muestran niveles de polimorfismo silencioso superiores. No obstante, tras la aplicación del test HKA no se detectaron diferencias significativas en los niveles de polimorfismo y divergencia de los genes *rp49* y *Dca*, lo que no permite concluir que haya un exceso de polimorfismo nucleotídico en el gen *Dca* de *D. subobscura*.

En *D. melanogaster* los valores de diversidad nucleotídica de la región *Dca* ($\pi_{\text{sil}} = 0.015$, $\pi_{\text{s}} = 0.026$) también se encuentran entre los más elevados detectados en la especie. La diversidad nucleotídica sinónima promedio de diferentes genes autosómicos en poblaciones no africanas de *D. melanogaster* es $\pi_{\text{s}} = 0.013$ (Moriyama y Powell 1996, Andolfatto 2001), aproximadamente la mitad que la detectada en el gen *Dca*. De hecho, únicamente el gen *Amy-p* presenta una diversidad nucleotídica sinónima comparable a la del gen *Dca*.

En la figura II.9 se muestra la distribución de la diversidad nucleotídica a lo largo de la región codificadora del gen *Dca* tanto en *D. subobscura* como en *D. melanogaster*. Como puede observarse, la ligeramente menor diversidad nucleotídica detectada en *D. melanogaster* en relación a *D. subobscura* se distribuye bastante homogéneamente a lo largo de toda la región analizada, excepto al inicio del exón 1. Además, se observa la presencia de regiones con picos altos de polimorfismo así como de zonas de baja o nula variabilidad que, por lo general, muestran una buena correspondencia entre las dos especies.

El primer pico de polimorfismo se ubica aproximadamente en el centro del primer exón con un valor máximo de $\pi = 0.023$ en *D. melanogaster* y de $\pi = 0.029$ en *D. subobscura*. El segundo pico se sitúa al final primer exón y presenta un valor de $\pi = 0.013$ en *D. melanogaster* y de $\pi = 0.032$ en *D. subobscura*. Finalmente, en el inicio del segundo exón (incluyendo el intrón) se localiza el pico con la diferencia más acusada de π entre las dos especies, con un valor de $\pi = 0.011$ en *D. melanogaster* y de $\pi = 0.042$ en *D. subobscura*.

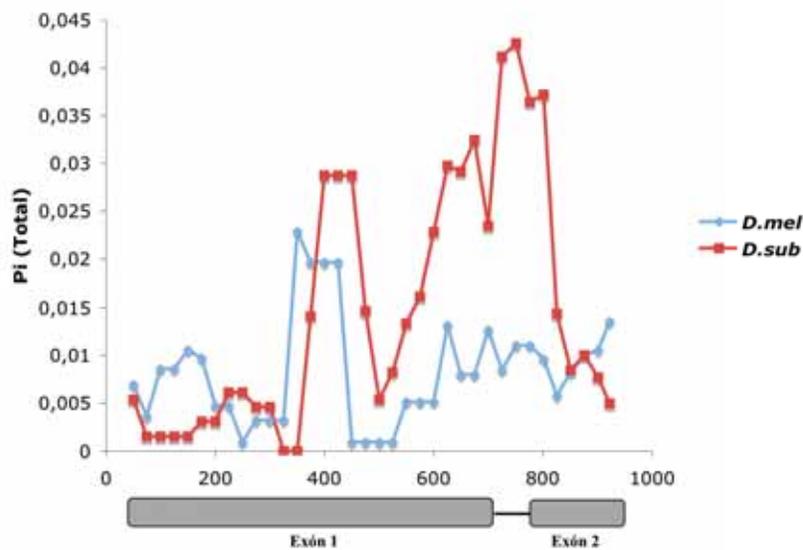


Figura II.9. Análisis de *sliding window* de la diversidad nucleotídica (π) a lo largo de la región codificadora del gen *Dca* en *D. subobscura* y *D. melanogaster*. Debajo de la gráfica se presentan los exones (en gris) y el intrón. Tamaño de la ventana de 100 pb con solapamientos de 25 pb.

El segundo pico de polimorfismo (localizado alrededor de la posición 700 en la figura II.9) incluye en *D. subobscura* diversos polimorfismos no sinónimos en fuerte desequilibrio de ligamiento. De hecho, tal como puede observarse en la figura II.7, las distintas variantes aminoacídicas presentes en las posiciones 731, 860, 959 y 976 segregan a frecuencias intermedias y se encuentran casi totalmente asociadas. Esta asociación se extiende además a otras posiciones sinónimas como la 951 y la 984. Este patrón no se detecta en *D. melanogaster* y contrasta con el patrón de polimorfismo que se detecta en la

zona 5' flanqueante de la región *Dca* en *D. subobscura* con un exceso significativo de variantes únicas.

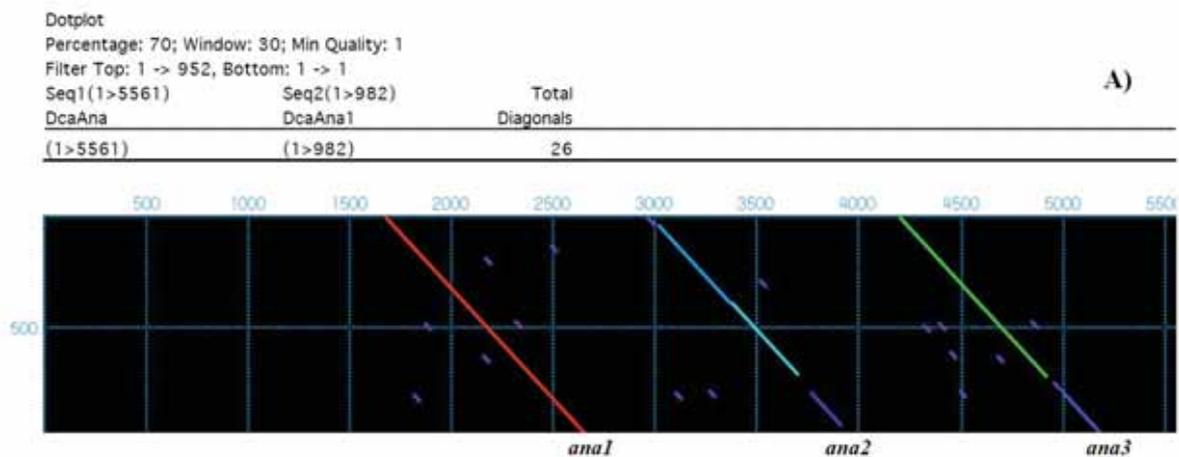
El patrón de polimorfismo en la región *Dca* en *D. melanogaster* y *D. subobscura* difiere considerablemente en las regiones flanqueantes. En ninguna de las especies el patrón de polimorfismo de la región 5' flanqueante concuerda con lo esperado bajo el modelo neutro tal y como indican los resultados significativos al aplicar distintos tests de neutralidad (tablas II.2 y II.5). No obstante, la naturaleza de la desviación detectada difiere en ambas especies. La región 5' flanqueante de *D. melanogaster* muestra una fuerte estructura haplotípica con un exceso de variantes a frecuencias intermedias y en fuerte desequilibrio de ligamiento entre ellas. Por el contrario, en *D. subobscura* la región 5' flanqueante está caracterizada por un exceso de variantes únicas. En la región 5' flanqueante del gen *Dca* de *D. melanogaster* se han descrito 13 elementos reguladores (Goto 2000), de los que 5 están incluidos en el fragmento analizado en este estudio. No obstante, únicamente uno de los elementos GATA-1 descritos presenta una posición polimórfica, lo que indica una fuerte restricción a variar de dichos elementos.

II.2 DIVERGENCIA DEL GEN *Dca*

II.2.1 RESULTADOS

II.2.1.1 DIVERGENCIA NUCLEOTÍDICA DEL GEN *Dca* EN *DROSOPHILA*

El gen *Dca* fue secuenciado en 5 especies del subgrupo *melanogaster* (*D. melanogaster*, *D. simulans*, *D. mauritiana*, *D. yakuba* y *D. erecta*) y 3 especies del grupo *obscura* de *Drosophila* (*D. subobscura*, *D. madeirensis* y *D. guanche*). Además, para realizar los diferentes análisis se utilizaron las secuencias nucleotídicas de la región codificadora del gen *Dca* de las especies *D. sechellia*, *D. ananassae*, *D. pseudoobscura* y *D. persimilis* obtenidas a partir de las bases de datos de los proyectos de secuenciación del genoma de estas especies. En *D. pseudoobscura* y *D. ananassae* se detectó más de una región con una elevada similitud con la región codificadora del gen *Dca*. El análisis de *Dot plot* confirmó el número de copias del gen presentes en estas dos especies. En *D. ananassae* se identificaron 3 copias que se encuentran formando un *cluster* en una región de aproximadamente 3.5 kb. Dos de las copias corresponden a genes funcionales (*ana1* y *ana3*) y la tercera puede considerarse pseudogen (*ana2*) debido a la presencia de un codón de *stop* prematuro, por lo que no fue empleada en la mayor parte de los análisis. En *D. pseudoobscura*, las dos regiones con alta similitud encontradas corresponden a sendas copias funcionales del gen (*pse1* y *pse2*) separadas por aproximadamente 10 kb (figura II.10).



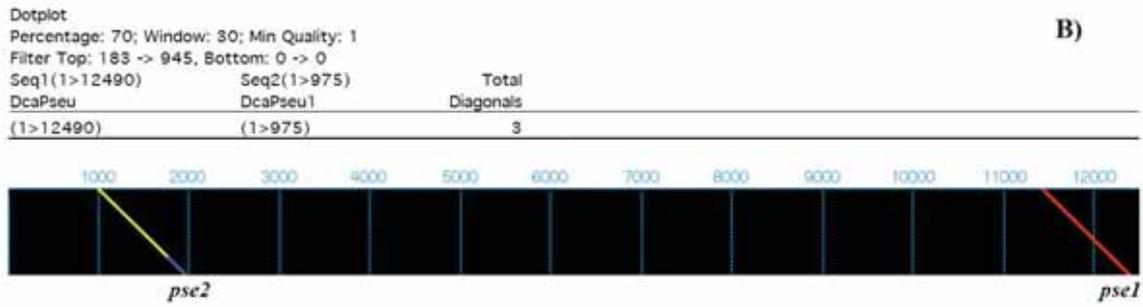


Figura II.10. Análisis de *Dot plot* del gen *Dca* en A) una región genómica de 5.5 kb de *D. ananassae* y B) una región genómica de 12.5 kb de *D. pseudoobscura*.

La longitud de la región codificadora de todas las especies analizadas es de 909 pb a excepción de *D. ananassae* la que difiere del resto de especies por una inserción de tres nucleótidos. Por lo tanto, el alineamiento múltiple de la región codificadora de las 14 secuencias consta de un total de 912 posiciones. En todas las especies, la estructura del gen *Dca* esta formada por dos exones separados por un pequeño intrón. La longitud de la región codificadora incluyendo el intrón varía desde 966 pb en *D. erecta* hasta 988 pb en *D. yakuba*.

La tabla II.7 muestra la matriz de distancias de las estimas de divergencia corregidas de acuerdo al método de Nei-Gojobori (1986) modificado por Nei y Kumar (Nei y Kumar 2000). Sobre la diagonal se muestra la divergencia sinónima (K_s) y bajo la diagonal la correspondiente divergencia no sinónima (K_a). En todas las comparaciones, las estimas de K_a son bastante inferiores a las correspondientes estimas de K_s .

Tabla II.7. Divergencia sinónima (K_s) y no sinónima (K_a) en el gen *Dca*

	<i>mel</i>	<i>sim</i>	<i>sec</i>	<i>mau</i>	<i>yak</i>	<i>ere</i>	<i>ana1</i>	<i>ana3</i>	<i>pse1</i>	<i>pse2</i>	<i>per</i>	<i>sub</i>	<i>mad</i>	<i>gua</i>
<i>mel</i>		0.139	0.094	0.134	0.183	0.157	0.703	0.702	0.642	0.652	0.639	0.608	0.639	0.618
<i>sim</i>	0.020		0.039	0.060	0.157	0.125	0.685	0.631	0.679	0.694	0.667	0.640	0.672	0.606
<i>sec</i>	0.025	0.011		0.052	0.133	0.102	0.679	0.648	0.642	0.647	0.622	0.606	0.637	0.599
<i>mau</i>	0.019	0.002	0.009		0.157	0.125	0.659	0.608	0.654	0.669	0.643	0.616	0.647	0.618
<i>yak</i>	0.039	0.035	0.043	0.033		0.096	0.646	0.609	0.576	0.598	0.566	0.560	0.589	0.573
<i>ere</i>	0.04	0.030	0.038	0.028	0.029		0.612	0.578	0.562	0.575	0.552	0.571	0.601	0.568
<i>ana1</i>	0.158	0.162	0.162	0.163	0.160	0.154		0.355	0.445	0.429	0.446	0.451	0.490	0.509
<i>ana3</i>	0.177	0.180	0.182	0.182	0.177	0.183	0.170		0.457	0.454	0.456	0.475	0.506	0.490
<i>pse1</i>	0.139	0.128	0.134	0.130	0.141	0.139	0.203	0.216		0.031	0.015	0.241	0.283	0.293
<i>pse2</i>	0.139	0.127	0.133	0.129	0.140	0.138	0.201	0.215	0.003		0.031	0.260	0.302	0.313
<i>per</i>	0.138	0.127	0.133	0.129	0.139	0.137	0.200	0.215	0.003	0		0.234	0.275	0.285
<i>sub</i>	0.16	0.146	0.151	0.147	0.150	0.156	0.201	0.217	0.048	0.051	0.051		0.050	0.100
<i>mad</i>	0.163	0.148	0.153	0.149	0.152	0.162	0.206	0.220	0.052	0.052	0.052	0.008		0.117
<i>gua</i>	0.158	0.144	0.149	0.145	0.149	0.154	0.207	0.215	0.050	0.051	0.051	0.014	0.009	

Estimas de divergencia de acuerdo al método de Nei-Gojobori (1986) modificado por Nei y Kumar (2000). Las estimas de divergencia sinónima (K_s) y no sinónima (K_a) se muestran sobre y bajo la diagonal, respectivamente. *mel*, *D. melanogaster*; *sim*, *D. simulans*; *sec*, *D. sechellia*; *mau*, *D. mauritiana*; *yak*, *D. yakuba*; *ere*, *D. erecta*; *ana*, *D. ananassae*; *pse*, *D. pseudoobscura*; *sub*, *D. subobscura*; *mad*, *D. madeirensis* y *gua*, *D. guanche*.

Las estimas de K_s entre las especies del subgrupo *melanogaster* y las incluidas en el grupo *obscura* oscilan entre 0.552 y 0.694. Estas estimas son considerablemente superiores a las detectadas entre *D. ananassae* y las especies del grupo *obscura*, con un rango que oscila entre 0.429 y 0.509 para *anal* y entre 0.454 y 0.506 para *ana3*. Por el contrario, las estimas de divergencia no sinónima (K_a) son mayores en las comparaciones entre las especies del grupo *obscura* y *D. ananassae* (desde 0.20 hasta 0.207 para *anal* y desde 0.215 hasta 0.220 para *ana3*) que en las comparaciones entre las especies del grupo *obscura* y las del subgrupo *melanogaster* (desde 0.127 hasta 0.163).

Para determinar si las diferencias en los niveles de divergencia sinónima detectados pueden estar relacionadas con diferencias en el *codon bias*, se estimó el sesgo en el uso de codones sinónimos para las diferentes especies (tabla II.8). Las especies del grupo *obscura* presentan niveles de *codon bias* superiores a los encontrados en las especies del subgrupo *melanogaster*. Los valores detectados en *D. ananassae* son similares a los que presentan las especies del grupo *obscura*, mientras que *D. yakuba* y *D. erecta* exhiben los niveles más altos de *codon bias* dentro del subgrupo *melanogaster*.

Tabla II.8. Estimaciones de *codon bias* en el gen *Dca*

Especies	Scaled χ^2	CBI	ENC
Subgrupo <i>melanogaster</i>			
<i>D. melanogaster</i>	0.363	0.425	44.315
<i>D. simulans</i>	0.387	0.455	42.122
<i>D. sechellia</i>	0.413	0.452	44.188
<i>D. mauritiana</i>	0.414	0.450	41.769
<i>D. yakuba</i>	0.433	0.483	40.975
<i>D. erecta</i>	0.456	0.489	39.639
Subgrupo <i>ananassae</i>			
<i>D. ananassae 1</i>	0.746	0.720	39.607
<i>D. ananassae 3</i>	0.694	0.691	37.531
Grupo <i>obscura</i>			
<i>D. pseudoobscura 1</i>	0.751	0.662	36.884
<i>D. pseudoobscura 2</i>	0.701	0.629	37.641
<i>D. persimilis</i>	0.740	0.653	36.385
<i>D. subobscura</i>	0.706	0.605	36.633
<i>D. madeirensis</i>	0.676	0.591	36.075
<i>D. guanche</i>	0.662	0.594	35.385

CBI, índice de *codon bias*; ENC, número efectivo de codones.

Los árboles obtenidos por el método de *neighbor-joining* a partir de las sustituciones sinónimas y no sinónimas se muestran en las figuras II.11A y II.11B, respectivamente. En ambos casos la mayoría de los nodos están soportados por porcentajes de *bootstrap* relativamente elevados. La topología observada es similar en los dos árboles, sin embargo en el obtenido en base a la divergencia sinónima la rama que lleva a las especies del subgrupo *melanogaster* es inusualmente larga. En el caso de la divergencia no sinónima destaca la longitud de las ramas que conducen a las diferentes copias del gen *Dca* de *D. ananassae*.

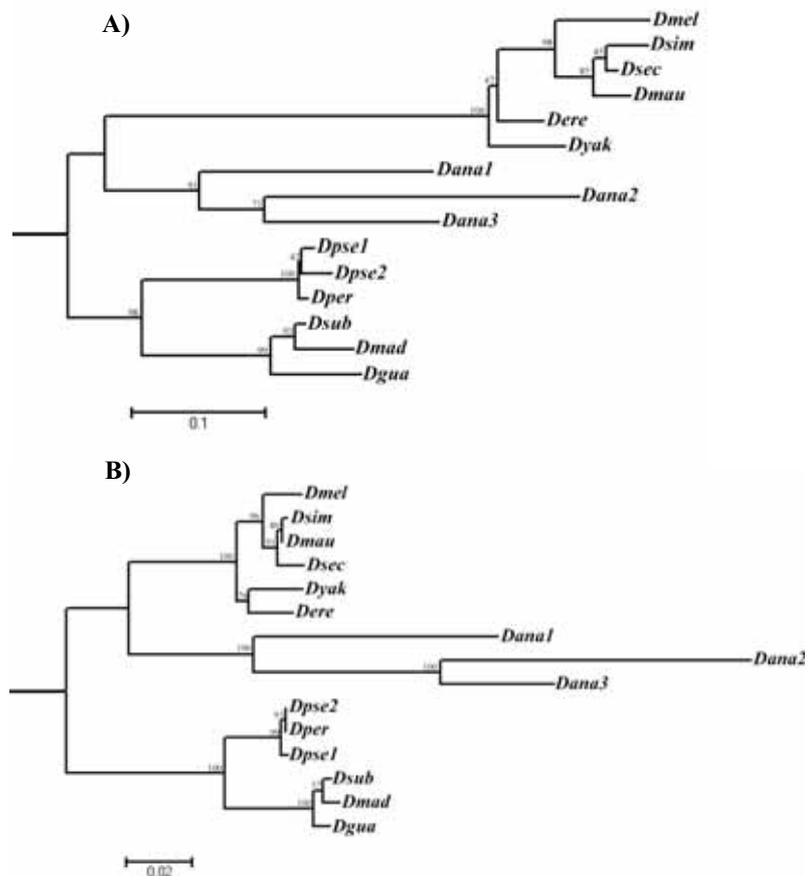


Figura II.11. Árboles filogenéticos obtenidos por el método de *neighbor-joining* (Saitou y Nei 1987) en base a la A) divergencia sinónima y B) no sinónima del gen *Dca*.

La aplicación del *relative-rate* test propuesto por Wu y Li (1985) para la divergencia total de la región codificadora del gen *Dca* entre las especies del grupo *melanogaster* no permitió detectar diferencias significativas en la tasa de sustitución nucleotídica. No obstante, al emplear el *relative-rate* test de Tajima (1993) para las sustituciones sinónimas y no sinónimas independientemente entre las especies del

subgrupo *melanogaster* y las tres copias del gen *Dca* en *D. ananassae* empleando *D. subobscura* como especie *outgroup* (tabla II.9), se detectó un exceso significativo de la tasa de fijación de las mutaciones no sinónimas en el linaje de *D. ananassae* en todas las comparaciones realizadas. Únicamente la comparación entre *ana1* y *mel* presenta desviaciones significativas de la constancia de la tasa de sustitución para las mutaciones sinónimas, siendo marginalmente significativas las comparaciones entre *ana1-sim*, *ana1-sec*, *ana1-mau* y *ana2-yak*.

Tabla II.9. Relative rate test de Tajima (1993)

Out- Group	Especie 1- Especie 2	Test de Tajima					
		Sustituciones sinónimas			Sustituciones no sinónimas		
		m1	m2	χ^2	r1	r2	χ^2
sub.....	<i>ana1-mel</i>	43	67	5.236*	42	17	10.593**
	<i>ana1-sim</i>	48	69	3.772 [#]	41	15	12.071**
	<i>ana1-sec</i>	49	66	2.513	40	17	9.281**
	<i>ana1-mau</i>	48	67	3.141 [#]	41	15	12.071**
	<i>ana1-yak</i>	48	61	1.551	41	16	10.964**
	<i>ana1-ere</i>	45	62	2.701	38	13	12.254**
sub	<i>ana2-mel</i>	62	53	0.704	63	12	34.681***
	<i>ana2-sim</i>	64	53	1.031	67	14	34.679***
	<i>ana2-sec</i>	66	52	1.661	66	15	32.111***
	<i>ana2-mau</i>	66	54	1.2	67	13	36.451***
	<i>ana2-yak</i>	71	51	3.278 [#]	63	12	34.680***
	<i>ana2-ere</i>	68	50	2.745	61	12	32.891***
sub.....	<i>ana3-mel</i>	55	64	0.680	41	20	7.229**
	<i>ana3-sim</i>	54	63	0.692	43	18	10.245**
	<i>ana3-sec</i>	59	62	0.007	40	20	6.667**
	<i>ana3-mau</i>	57	62	0.210	41	18	8.966**
	<i>ana3-yak</i>	59	59	1	40	17	9.281**
	<i>ana3-ere</i>	55	58	0.079	40	18	8.344**

m1 y m2, número de sustituciones sinónimas en los linajes que llevan a las especies 1 y 2, respectivamente.

r1 y r2, número de sustituciones no sinónimas en los linajes que llevan a las especies 1 y 2, respectivamente.

Nombre de las especies como en la tabla II.7.

* $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$

Evolución del intrón

La estructura del gen *Dca* en todas las especies analizadas se encuentra formada por dos exones separados por un intrón que puede ser clasificado de fase 2 puesto que interrumpe la pauta de lectura del gen entre la segunda y la tercera base de un codón.

La posible diferencia en los patrones de sustitución nucleotídica, inserciones y deleciones en el intrón, fue analizada dentro de las especies del *cluster melanogaster* (*D. melanogaster*, *D. simulans*, *D. sechellia* y *D. mauritiana*) y entre aquellas que conforman el *cluster subobscura* (*D. pseudoobscura*, *D. subobscura*, *D. madeirensis* y *D. guanche*), independientemente.

Dentro del subgrupo *melanogaster*, *D. yakuba* presenta el intrón más largo con un tamaño 76 pb, mientras que en el grupo *obscura* es *D. persimilis* la especie con el mayor intrón (62 pb). Por lo tanto, el intrón presente en todas las especies analizadas puede catalogarse dentro de la clase de intrones cortos (Mount et al. 1992; Yu et al. 2002). El tamaño del intrón en cada especie, así como las estimas de la divergencia en el subgrupo *melanogaster* y el grupo *obscura* se muestran en la tabla II.10.

Tabla II.10. Tamaño del intrón en el gen *Dca*

Especies	Tamaño	Divergencia intrón
Subgrupo <i>melanogaster</i>		
<i>mel-sim</i>	58-56 (58)	0.118
<i>mel-sech</i>	58-56 (58)	0.162
<i>mel-mau</i>	58-56 (58)	0.140
<i>mel-yak</i>	58-76 (78)	0.388
<i>mel-ere</i>	58-55 (58)	0.314
<i>sim-sech</i>	56-56 (56)	0.077
<i>sim-mau</i>	56-56 (56)	0.019
<i>sim-yak</i>	56-76 (76)	0.274
<i>sim-ere</i>	56-55 (56)	0.211
<i>sech-mau</i>	56-56 (56)	0.097
<i>sech-yak</i>	56-76 (76)	0.291
<i>sech-ere</i>	56-55 (56)	0.258
<i>mau-yak</i>	56-76 (76)	0.306
<i>mau-ere</i>	56-55 (56)	0.186
<i>yak-ere</i>	76-55 (76)	0.317
Grupo <i>obscura</i>		
<i>per-sub</i>	62-59 (66)	0.398
<i>per-made</i>	62-59 (66)	0.365
<i>per-guan</i>	62-58 (66)	0.370
<i>sub-made</i>	59-59 (59)	0.101
<i>sub-guan</i>	59-58 (59)	0.145
<i>made-guan</i>	59-58 (59)	0.123

Nombre de las especies como en la tabla II.7. Entre paréntesis, longitud del alineamiento entre pares de especies.

De acuerdo al criterio de máxima parsimonia se infirió el número de sustituciones nucleotídicas y de *indels* que presenta la secuencia del intrón en las especies del *cluster melanogaster*. Con esta finalidad se utilizaron *D. yakuba* y *D. erecta* como especies *outgroup*. La especie empleada como *outgroup* en el *cluster*

subobscura fue *D. persimilis* (tabla II.11). En el *cluster melanogaster* se detectaron un total de 19 sustituciones y 2 *indels* en la comparación con *D. yakuba* y un total de 17 sustituciones y 2 *indels* en la comparación con *D. erecta*.

Tabla II.11. Número de sustituciones e indels en el intrón del gen *Dca*

Especies	Alineamiento	Sub	Sub/pb ^a	Indels	Indels/pb	Del	Ins	Del/Ins
Cluster <i>melanogaster</i>								
<i>D. yakuba</i> ^b	78	19	0.243	2	0.025	1	1	1
<i>D. erecta</i> ^b	58	17	0.293	2	0.034	0	2	-
Cluster <i>obscura</i>								
<i>D. persimilis</i> ^c	68	18	0.264	5	0.073	3	2	1.5

Sub, sustituciones; Del, deleciones; Ins, inserciones.

^a Número total de pares de bases (incluyendo los *gaps*) en el alineamiento.

^b Especies empleadas como *outgroup* para el *cluster melanogaster*.

^c Especies empleadas como *outgroup* para el *cluster obscura*.

El *cluster subobscura* presenta 18 sustituciones y 5 *indels*. La relación sustituciones/*indels* para el *cluster melanogaster* es muy similar en las dos comparaciones (9.5 al usar *D. yakuba* como *outgroup* y 8.5 al usar *D. erecta*), mientras que en el *cluster subobscura* la relación observada es inferior (3.6). Sin embargo, el test de χ^2 de contingencia realizado a partir de los valores observados de sustituciones nucleotídicas y de *indels* detectados en cada *cluster* no es significativo ($\chi^2 = 1.224$; $P = 0.268$ y $\chi^2 = 0.942$; $P = 0.331$ al utilizar *D. yakuba* o *D. erecta* como *outgroups*, respectivamente). El número de sustituciones por posición es muy similar en los dos *cluster*, con valores de 0.243 sustituciones/pb en la comparación con *D. yakuba* y de 0.293 sustituciones/pb en la comparación con *D. erecta* dentro del *cluster melanogaster* y un valor de 0.264 sustituciones/pb en el *cluster subobscura*.

Los *indels* detectados fueron clasificados como inserciones o deleciones. De los dos *indels* hallados en el *cluster melanogaster* en la comparación con *D. yakuba*, uno es una inserción de 2 pb y el otro es una deleción de 20 pb. En la comparación con *D. erecta* también se detectaron 2 *indels* en este *cluster*, pero en este caso ambos son inserciones de 2 pb y 1pb. En el *cluster subobscura* se identificaron 5 *indels*, de los cuales 3 representan deleciones de 4 pb, 1pb y 5 pb y los dos restantes son inserciones de 5 pb y 2 pb.

La relación deleciones/inserciones es de 1 para el *cluster melanogaster* en la comparación con *D. yakuba* y de 1.5 en el *cluster subobscura*. Estas estimas son muy parecidas al valor de la relación entre deleciones/inserciones de 1.35 detectado por

Comeron y Kreitman (2000) en un estudio de polimorfismo de *indels* en intrones de *D. melanogaster* y por Parsch (2003) en un estudio realizado en el subgrupo *melanogaster*.

Genes parálogos

Se detectó la presencia de más de una región con una alta similitud al gen *Dca* en *D. pseudoobscura* y *D. ananassae* (figura II.10). En la primera especie, se identificaron dos regiones que corresponden a genes funcionales y se encuentran separadas aproximadamente 10 kb. En *D. ananassae*, fueron detectadas tres regiones que se encuentran formando un *cluster* de aproximadamente 3.5 kb. Dos de estas regiones (*ana1* y *ana3*) corresponden a genes funcionales mientras que la tercera (*ana2*) puede considerarse pseudogen puesto que presenta un codón de *stop* prematuro en la posición 90 de la secuencia aminoacídica (figura II.12A).

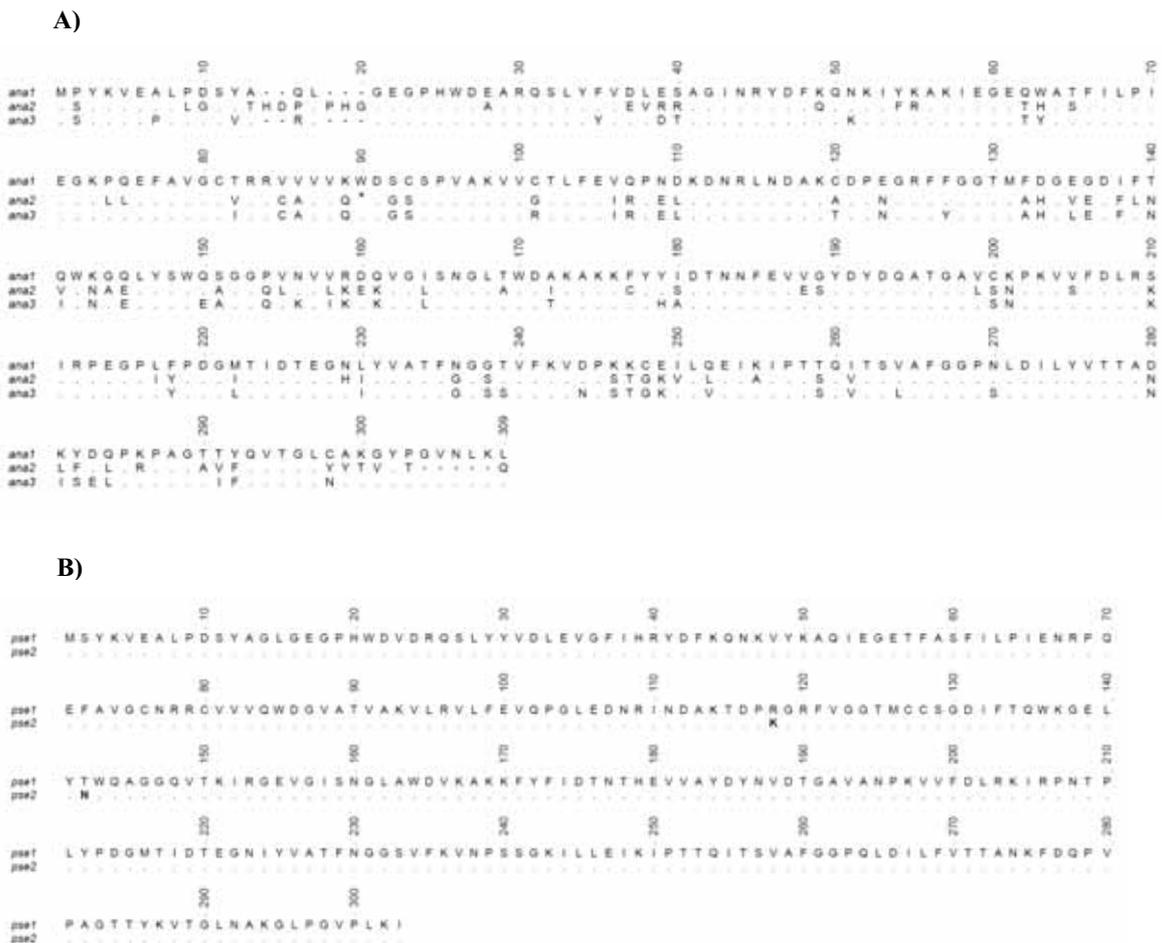


Figura II.12 Alineamiento múltiple de la secuencia aminoacídica de A) las tres copias del gen *Dca* detectadas en *Drosophila ananassae* y B) de las dos copias encontradas en *Drosophila pseudoobscura*. Los puntos (.) indican aminoácidos idénticos a los de la secuencia de referencia. Los guiones (-) deleciones de aminoácidos. El asterisco (*) muestra el codón de *stop* prematuro en la copia *ana2*.

El alineamiento múltiple de la secuencia aminoacídica codificada por los 3 genes parálogos en *D. ananassae* comprende un total de 309 aminoácidos (figura II.12A). En *ana2* se detectaron dos inserciones y una deleción con respecto a las copias *ana1* y *ana3*. Las dos inserciones se encuentran en el inicio de la región codificadora y comprenden un total de 2 y 3 aminoácidos, mientras que la deleción se ubica al final del gen y abarca 5 aminoácidos. Por lo tanto, las tres copias presentan una pauta de lectura de igual longitud (304 aminoácidos). La comparación *ana1-ana2* presenta 99 diferencias sinónimas y 116 no sinónimas, mientras que en *ana1-ana3* las diferencias sinónimas se reducen a 79 y las no sinónimas a 90. En la comparación *ana2-ana3* las diferencias sinónimas y no sinónimas observadas fueron 81 y 71, respectivamente. En *D. pseudoobscura*, el alineamiento de la secuencia aminoacídica codificada por las dos copias del gen *Dca* comprende 303 aminoácidos (figura II.12B). El número de diferencias nucleotídicas detectadas entre las dos copias es de 10 (8 sinónimas y 2 no sinónimas). No se observaron deleciones ni inserciones.

Las estimas de la divergencia nucleotídica corregidas por el método de Nei-Gojobori (1986) modificado (Nei y Kumar 2000) entre los diferentes genes parálogos de *D. ananassae* y de *D. pseudoobscura* se muestran en la tabla II.12. Las estimas de divergencia silenciosa corresponden a la divergencia en las posiciones sinónimas y en el intrón. En *D. ananassae*, la comparación *ana1-ana2* presenta las estimas de divergencia más elevadas para todos los tipos de posiciones. Los menores valores tanto de divergencia sinónima como silenciosa se encuentran en la comparación *ana1-ana3*, mientras que la comparación *ana2-ana3* presenta los niveles más bajos de divergencia no sinónima. Todas las estimas de divergencia en la comparación entre *pseu1-pseu2* son muy bajas, especialmente para las posiciones no sinónimas.

Tabla II.12. Divergencia nucleotídica entre genes parálogos

Especie	Sinónima	No. posiciones sinónimas	No sinónima	No. posiciones no sinónimas	silenciosa	No. posiciones silenciosas	K_s / K_a
<i>ana1 - ana2</i>	0.459	216.5	0.228	680.5	0.609	273.5	0.500
<i>ana1 - ana3</i>	0.355	219.417	0.170	692.583	0.467	285.417	0.478
<i>ana2 - ana3</i>	0.364	222.75	0.129	674.25	0.504	280.75	0.354
<i>pse1 - pse2</i>	0.031	223.833	0.003	685.167	0.032	284.833	0.097

Nombre de las especies como en la tabla II.7.

La relación entre la divergencia no sinónima (K_a) y sinónima (K_s) entre las diferentes copias del gen *Dca* en *D. ananassae* y en *D. pseudoobscura* son menores de 1. Sin embargo, el valor observado varía desde 0.354 en la comparación *ana2-ana3* hasta 0.5 en la de *ana1-ana2*, mientras que en *pseu1-pseu2* este valor es

extremadamente bajo (0.097) (tabla II.12). En la figura II.13 se muestra la distribución de la relación K_a/K_s a lo largo de la región codificadora entre las diferentes comparaciones de *D. ananassae*. A pesar del relativamente buen ajuste en la distribución del ratio K_a/K_s en las distintas comparaciones, la zona central del gen contiene regiones con valores de $K_a/K_s > 1$.

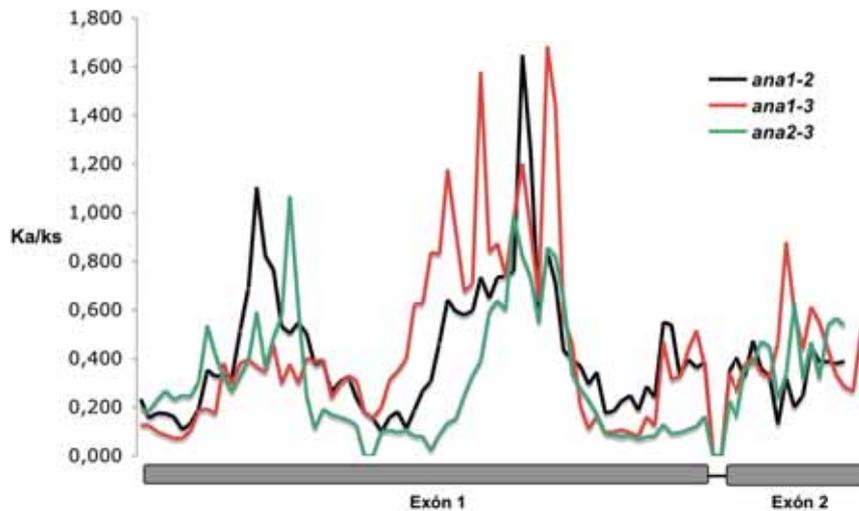


Figura II.13. Análisis de *sliding window* de la relación entre la divergencia no sinónima (K_a) y sinónima (K_s) en las diferentes comparaciones entre las copias del gen *Dca* en *D. ananassae*. El tamaño de la ventana es de 70 posiciones con desplazamientos de 10 posiciones.

Análisis de *maximum likelihood* (ML) de la región codificadora del gen *Dca*

La evolución molecular y la divergencia de los genes *Dca* parálogos y ortólogos se analizaron usando diferentes modelos evolutivos implementados en el programa PAML (Yang 1997) y de acuerdo a la filogenia de la figura II.14. El análisis fue realizado en 12 especies del subgénero *Sophophora*, aunque el número total de secuencias utilizadas fue de 14 al incluirse las dos copias funcionales detectadas en *D. ananassae* y *D. pseudoobscura*. Las secuencias de los genes *anterior fat body protein* (AFP) de *Calliphora vicina* (*Blue Blowfly*) y *Sarcophoga peregrina* (*Fleshfly*) con elevada similitud al gen *Dca* se utilizaron como *outgroup*.

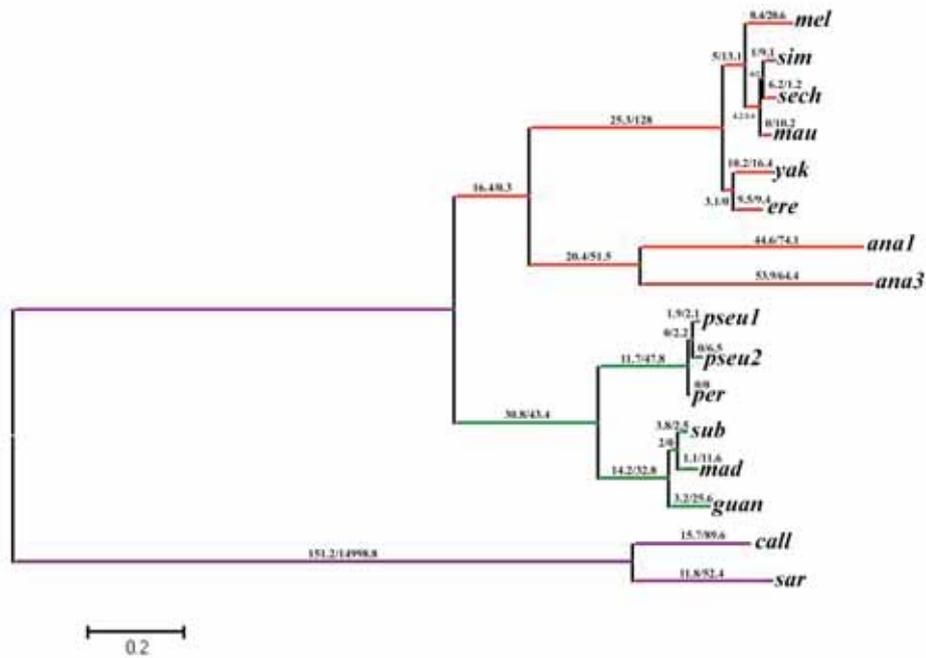


Figura II.14 Filogenia de las 14 secuencias del gen *Dca* en las 12 especies analizadas de *Drosophila* y las dos especies empleadas como *outgroup*. *Call*, *Calliphora vicina*; *sar*, *Sarcophoga peregrina*. Los números sobre cada una de las ramas son las estimas de *maximum likelihood* del número de sustituciones no sinónimas/sinónimas para cada rama empleando el modelo FR. La longitud de las ramas se muestra en proporción a las estimas de la divergencia. En rojo, ramas tropicales; en verde, ramas templadas; en lila el resto de ramas (*outgroup*)

Los valores de *log likelihood* (ℓ) así como las estimas de los parámetros bajo diferentes modelos se muestran en la tabla II.13. El modelo M0 (*one-ratio model*) asume la misma relación d_N/d_S ($d_N/d_S = \omega$) para todas las ramas de la filogenia y presenta un $\ell_0 = -5405.100$. Por el contrario, de acuerdo al modelo FR (*free-ratio model*) que admite una ω diferente para cada una de las ramas del árbol, infiere un valor de $\ell_1 = -5356.091$. Puesto que el modelo M0 incluye 31 parámetros y el modelo FR 59, dos veces la diferencia de *log likelihood*, $2\Delta\ell = 2(\ell_1 - \ell_0) = 98.017$ puede ser comparada con una distribución χ^2 con 28 grados de libertad para poder determinar si el modelo FR se ajusta mejor a los datos que el modelo M0. La diferencia entre los dos modelos es significativa ($P = 1.058 \times 10^{-9}$), indicando que el valor de ω es diferente en los distintos linajes.

Tabla II.13. Valores de *log likelihood* y estimas de los parámetros bajo diferentes modelos

Modelo	f	ℓ	Estimas de los parámetros	Selección positiva
Branch models				
M0	31	- 5405.100666	$\omega = 0.089, \kappa = 1.41$	No
FR	59	- 5356.091978	$\kappa = 1.32$	No
BrTrop-Temp	33	- 5387.774742	$\omega_{Trop} = 0.124, \omega_{Temp} = 0.082$ $\omega_{Res} = 0.029$ $\kappa = 1.32$	No
Sites models				
M1	32	- 5329.045542	$\kappa = 1.57$ $p_0 = 0.873, \omega_0 = 0.060$ $p_1 = 0.126, \omega_1 = 1$	No permitido
M2	34	- 5329.045542	$\kappa = 1.57$ $p_0 = 0.873, \omega_0 = 0.060$ $p_1 = 0.126, \omega_1 = 1$ $p_2 = 0.000, \omega_2 = 28.251$	No
M3 (K=2)	33	- 5289.371589	$\kappa = 1.44$ $p_0 = 0.716, \omega_0 = 0.027$ $p_1 = 0.283, \omega_1 = 0.307$	No
M3 (K=3)	35	- 5289.371589	$\kappa = 1.44$ $p_0 = 0.716, \omega_0 = 0.027$ $p_1 = 0.283, \omega_1 = 0.307$ $p_2 = 0.000, \omega_2 = 20.96$	No
M7	32	- 5283.764578	$\kappa = 1.43$ $p = 0.387, q = 2.974$	No permitido
M8	34	- 5279.324126	$\kappa = 1.44$ $p = 0.467, q = 4.481$ $p_0 = 0.977, p_1 = 0.022$ $\omega_1 = 1.281$	No
Branch-sites model A				
RamasTrop	34	- 5318.332077	$\kappa = 1.53$ $p_0 = 0.831, p_1 = 0.075$ $(p_2 + p_3) = 0.093$ $\omega_2 = 1.000$	104D ($P > 0.95$)
RamasTemp	34	- 5320.676155	$\kappa = 1.51$ $p_0 = 0.872, p_1 = 0.0879$ $(p_2 + p_3) = 0.039$ $\omega_2 = 2.216$	140W, 146S ($P > 0.95$)

f, grados de libertad; ℓ , valor de *log likelihood*; ω , relación d_N/d_S ; κ , relación entre el número de transiciones y transversiones; Trop, ramas tropicales; Temp, ramas templadas; Res, resto de ramas.

Las estimas de los valores de ω detectados en las diferentes ramas de acuerdo al modelo FR indican que el gen *Dca* presenta unos niveles de restricción funcional elevados. El número de sustituciones sinónimas y no sinónimas para cada una de las ramas del árbol fueron calculadas a partir de los parámetros estimados bajo el modelo FR y se presentan en la figura II.14.

Un factor que podría contribuir a dicha heterogeneidad es la diferencia en la distribución geográfica de las especies estudiadas. Por este motivo, la divergencia entre ellas se analizó atendiendo a las condiciones climáticas de su área de distribución. Con esta finalidad, estas 12 especies se han clasificado en templadas o tropicales de acuerdo a Gilbert et al (2001). Estos autores detectaron una muy buena relación entre la distribución de diversas especies de *Drosophila* y su capacidad de recuperación tras un tratamiento prolongado a 0°C. Las especies templadas se recuperaban mucho más rápidamente que las especies tropicales. Este resultado les permitió caracterizar, por ejemplo, a *D. melanogaster* y *D. simulans* como especies tropicales de acuerdo a su origen climático y a pesar de presentar en la actualidad una distribución cosmopolita. En base a esta idea, *D. madeirensis* y *D. guanche* se han considerado especies templadas aunque su distribución actual es subtropical.

Para determinar la posible diferencia en la evolución del gen *Dca* en linajes específicos causada por la asociación con la distribución geográfica, se empleó un modelo (*BrTrop-Temp*) con tres valores de ω diferentes: uno (ω_{Trop}) para los linajes tropicales, uno (ω_{Temp}) para los linajes templados y otro (ω_{Res}) para el resto de ramas (*outgroup*). Este modelo es significativamente mejor que el modelo M0 ($2\Delta\ell = 34.651848$, $P = 2.988 \times 10^{-8}$), indicando que el valor de ω para las ramas tropicales ($\omega_{\text{Trop}} = 0.124$) es significativamente diferente al obtenido para las ramas templadas ($\omega_{\text{Temp}} = 0.082$). Un resultado similar se obtuvo si *D. melanogaster* y *D. simulans* se consideraban especies templadas o si *D. madeirensis* y *D. guanche* se consideraban tropicales. El modelo *BrTrop-Temp* también se comparó con el modelo FR. En este caso, el modelo FR se ajusta mejor a los datos ($2\Delta\ell = 63.365528$, $P = 5.821 \times 10^{-5}$), indicando que otros factores además del carácter tropical o templado de las especies está contribuyendo a la heterogeneidad de los valores de ω entre las ramas.

De acuerdo con los anteriores modelos, el parámetro ω se estima como un promedio sobre todas las posiciones de las secuencias analizadas. Sin embargo, este criterio es muy restrictivo para poder detectar selección positiva puesto que la mayoría de los cambios adaptativos normalmente afectan a pocos aminoácidos de la proteína mientras que los restantes permanecen muy conservados. Por lo tanto, el valor de ω calculado a partir del criterio anterior puede no ser significativamente >1 incluso si se ha producido algún evento de adaptación molecular. Para evitar este problema y poder identificar posibles aminoácidos bajo selección positiva, se emplearon modelos

estadísticos que permiten que el parámetro ω varíe entre las diferentes posiciones (*site models*) (Nielsen y Yang 1998; Yang et al. 2000).

Los resultados de las estimas de *ML* al emplear los modelos M1 (*nearly neutral*), M2 (selección), M3 (discreto), M7 (*beta*) y M8 (*beta & ω*) se muestran en la tabla II.13. Los análisis realizados con los modelos M1, M2 y M3 ($K=2$ y $K=3$) indican que el 71-87% de las posiciones analizadas se encuentran bajo una fuerte selección purificadora con valores de ω que van desde 0.03 hasta 0.06. Sin embargo, de acuerdo al modelo M3 ($K=2$) en el resto de posiciones la selección es más relajada con valores de $\omega = 0.31$ (en los modelos M1 y M2 se fija una clase de ω con valor 1). Las estimas de los parámetros obtenidos con los modelos M7 y M8 sugieren que los valores de ω presentan una distribución *beta* en forma de *L*, con la mayoría de aminoácidos muy conservados o poco variables ($p \sim 0.4$ y $q \sim 3.5$).

Los modelos M1 y M2 fueron comparados para determinar si existen posiciones bajo selección natural. Ambos modelos no difieren en las estimas de *ML* por lo que no existen evidencias de selección natural en ningún codón en particular. Sin embargo, los modelos discretos (M3 con $K = 2$ o $K = 3$) se ajustan significativamente mejor que el modelo M0 a los datos. No obstante, los parámetros estimados bajo el M3 ($K = 3$) indican que ningún sitio se encuentra bajo selección positiva. Adicionalmente, este modelo no es significativamente mejor que el modelo más simple M3 ($K = 2$) proporcionando evidencia de la no existencia de sitios con $\omega > 1$. Sin embargo, existen dos categorías de posiciones evolucionando con diferentes niveles de restricción (M3($K = 2$), $\omega_0 = 0.027$ y $\omega_1 = 0.307$). Similarmente, a pesar de que el modelo M8 identifica una pequeña proporción de sitios ($p_1 = 2.2\%$) con un valor de $\omega = 1.28$, no identifica ninguna posición bajo selección positiva

El modelo alternativo A de *branch-sites* fue implementado con el propósito de determinar la existencia de sitios bajo selección positiva en linajes específicos. Este modelo, el cual es una extensión del modelo M1, asume que hay dos clases de sitios ($\omega_0 < 1$ y $\omega_1 = 1$) en todos los linajes y permite que algunos sitios evolucionen bajo selección positiva, incorporando una $\omega_2 > 1$, en el linaje de interés (*foreground branches*). Para los análisis se escogieron dos tipos de ramas como *foreground*: las que llevan a los linajes tropicales y a las que llevan a los linajes templados. Este modelo se ajusta significativamente mejor que el modelo M1 en ambos casos (tropicales, $2\Delta\ell = 21.426$, $P = 2.224 \times 10^{-5}$; templados, $2\Delta\ell = 16.738$, $P = 0.000232$), e identifica 1 (104D)

y 2 (140W, 146S) (figura II.15) posiciones bajo selección positiva, respectivamente. Por último, para evitar la identificación de falsos positivos con el modelo de *branch-sites*, se empleó el test 2 (Zhang et al. 2005) que compara el modelo alternativo A y el modelo nulo A con $\omega_2 = 1$. No obstante, este test es significativamente mejor únicamente en las ramas templadas. Por lo tanto, a pesar de que los modelos de *branch-sites* predicen posiciones seleccionadas positivamente, no se puede descartar que el valor de $\omega_2 > 1$ en los linajes tropicales sea causado por la relajación de la restricción funcional ($\omega_2 \approx 1$), mientras que la selección positiva puede haber actuado en las ramas templadas.

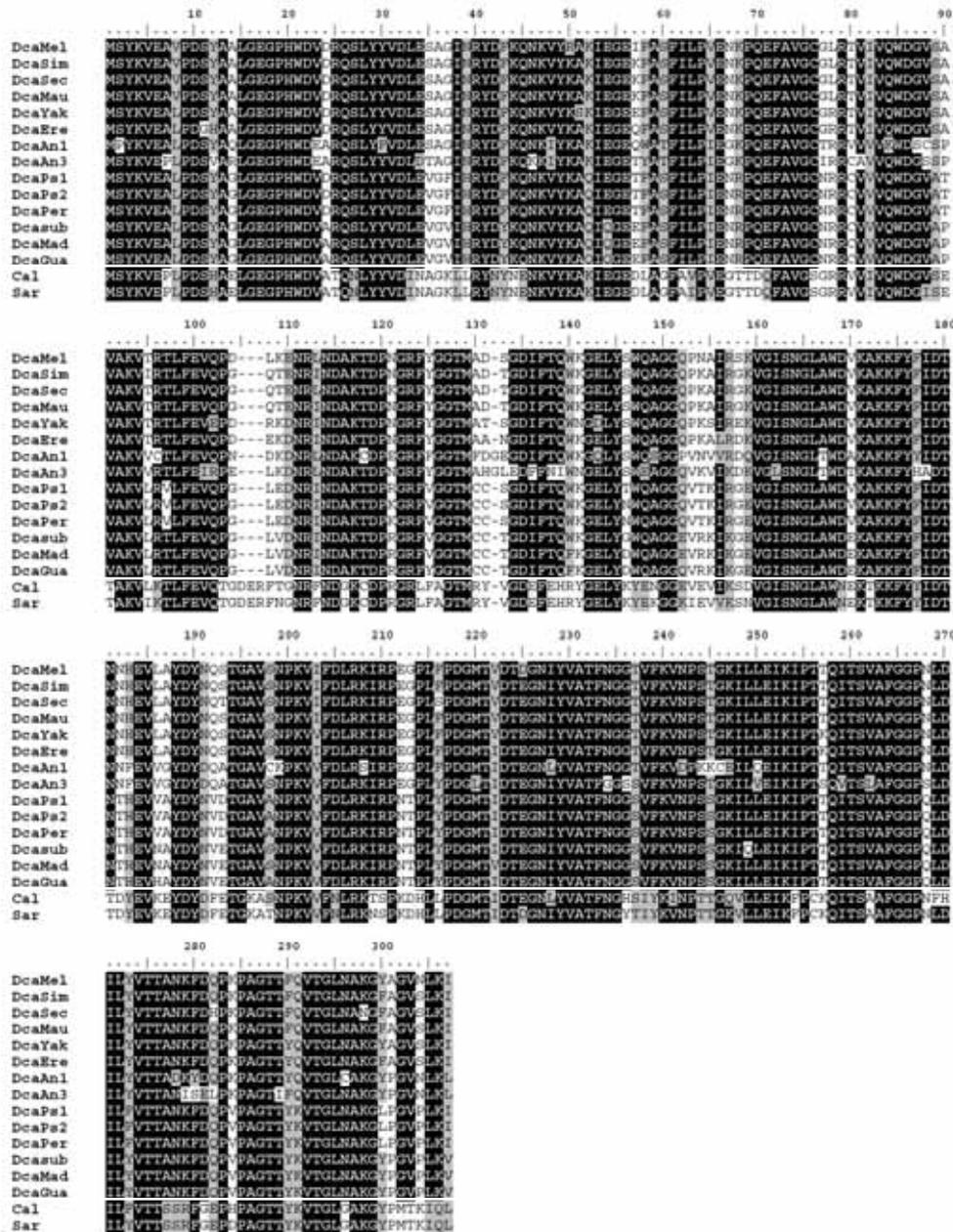


Figura II.15. Alineamiento múltiple de la proteína codificada por el gen *Dca* en las especies analizadas. En negro se indican aminoácidos idénticos en todas las secuencias y en gris aminoácidos con propiedades bioquímicas similares. Mel, *D. melanogaster*; Sim, *D. simulans*; Sec, *D. sechelia*; Mau, *D. mauritiana*; Yak, *D. yakuba*; Ere, *D. erecta*; Ana, *D. ananassae*; Ps, *D. pseudoobscura*; Per, *D. persimilis*; Sub, *D. subobscura*; Mad, *D. madeirensis*; Gua, *D. guanche*; Cal, *Calliphora vicina*; Sar, *Sarcophaga peregrina*.

II.2.2 DISCUSIÓN

II.2.2.1 DIVERGENCIA NUCLEOTÍDICA DEL GEN *DCA* EN *DROSOPHILA*

Los resultados obtenidos muestran que las estimas de la divergencia sinónima (K_s) entre las especies del grupo *obscura* y *D. ananassae* fueron inferiores que las detectadas entre las especies del grupo *obscura* y las incluidas en el subgrupo *melanogaster* a pesar que el tiempo de divergencia coincide en ambas comparaciones. Esta diferencia parecería indicar una aceleración en la tasa de fijación de mutaciones sinónimas en el linaje que conduce a las especies del subgrupo *melanogaster*. Esta aceleración también se ve reflejada en el árbol filogenético de la figura II.11A en el que destaca la longitud de la rama que conduce a dichas especies. Estos resultados pueden relacionarse con las diferencias en el sesgo en el uso de codones sinónimos entre las especies. De hecho, las especies del grupo *obscura* y *D. ananassae* muestran un sesgo en el uso de codones sinónimos más acusado que las especies del subgrupo *melanogaster*. El mantenimiento en el sesgo en el uso de codones sinónimos en *Drosophila* puede explicarse por la acción de una selección débil actuando en contra de las mutaciones de codón preferente a no preferente que podrían considerarse ligeramente deletéreas y a favor de las mutaciones de codón no preferente a preferente que serían ligeramente ventajosas (Akashi 1995).

Así, la aceleración significativa observada por medio del *relative rate test* en la tasa de fijación de las mutaciones sinónimas en el linaje de *D. melanogaster* y parcialmente significativa en el resto de especies del cluster *melanogaster*, podría deberse a una relajación de la selección débil en contra de las mutaciones de codón preferente a no preferente lo que provocaría tanto un aumento en la tasa de sustitución sinónima y una disminución del sesgo en el uso de codones sinónimos. Esta explicación también estaría de acuerdo con los datos del polimorfismo en *D. melanogaster*. De hecho, utilizando *D. simulans* como especie *outgroup* puede inferirse que de los 16 polimorfismos sinónimos detectados en la región codificadora del gen *Dca* en *D. melanogaster*, 10 producen un cambio de codón preferente a no preferente, mientras que tan solo 3 producen un cambio de codón no preferente a preferente. Además 5 de las 10 mutaciones de codón preferente a no preferente presentan una frecuencia superior al 50% indicando que están segregando a frecuencia elevada.

Se ha propuesto que en las regiones con niveles de recombinación reducidos la tasa de fijación de mutaciones ligeramente deletéreas es mayor que en las regiones con recombinación normal (Birky y Walsh 1988; Charlesworth 1994). Esta predicción se ha comprobado empíricamente ya que en *D. melanogaster* los genes localizados en zonas de baja recombinación tienden a presentar un bajo sesgo en el uso de codones sinónimos (Kliman y Hey 1993). En el estudio interespecífico realizado por Munté et al. (2001) de la divergencia del gen *yellow* (*y*) empleando las mismas especies analizadas en este trabajo se obtuvo un resultado similar al detectado para el gen *Dca*. Los resultados del gen *y* se interpretaron en relación al fuerte cambio recombinacional que exhibe la región *y* en estas especies ya que dicha región está localizada en una zona con una fuerte reducción de la recombinación en las especies del subgrupo *melanogaster* y de recombinación normal en *D. ananassae* y las especies del grupo *obscura*. Sin embargo, esta explicación no puede aplicarse para el gen *Dca* ya que éste no se encuentra ubicado en una región con un nivel reducido de recombinación en ninguna de las especies analizadas.

Por otra parte, la teoría casi neutral (Otha 1992) propone que en los linajes con tamaños efectivos pequeños se espera que la tasa de fijación de mutaciones ligeramente deletéreas sea más alta debido a que la deriva genética contrarresta el efecto de la selección débil. Así, los resultados de la divergencia interespecífica del gen *Dca* podrían explicarse asumiendo que el tamaño efectivo de las especies del subgrupo *melanogaster* es menor que el de *D. ananassae* y el de las especies que conforman el grupo *obscura*. Adicionalmente, dentro del grupo *melanogaster* los valores negativos de la diferencia $K_{13}-K_{23}$ en todas las comparaciones entre los linajes de *D. yakuba* o *D. erecta* con los del resto de especies del subgrupo *melanogaster* al utilizar *D. ananassae* como *outgroup* indican una tasa de sustitución sinónima mayor en los linajes que llevan a *D. melanogaster*, *D. simulans*, *D. sechelia* y *D. mauritiana*. Este resultado también sería consistente con una mayor eficacia de la selección en contra de las mutaciones ligeramente deletéreas en los linajes de *D. yakuba* y *D. erecta* debido a un mayor tamaño efectivo. De hecho, estas especies presentan los niveles de codon bias más elevados entre las especies del subgrupo *melanogaster*.

Por otro lado, la aplicación del *relative rate test* permitió detectar una aceleración significativa en la tasa de fijación de mutaciones no sinónimas en las diferentes copias del gen *Dca* en *D. ananassae* (tabla II.9). Este resultado puede ser causado por la relajación de las constricciones funcionales que permitiría la

acumulación de sustituciones no sinónimas a lo largo del gen. No obstante, la tasa de fijación de mutaciones sinónimas y no sinónimas en las comparaciones que incluyen *ana2* son las más elevadas lo que concuerda con la idea que este gen se encuentra en proceso de pseudogenización.

Genes parálogos

En este estudio fueron identificadas 3 regiones de alta similitud a la región codificadora del gen *Dca* en *D. ananassae* y 2 en *D. pseudoobscura*. De las 3 regiones de la primera especie, una fue considerada como un pseudogen debido a la presencia de un codón de *stop* en la posición 90 del alineamiento (figura II.12A). Sin embargo, ninguna otra característica de pseudogenización como la presencia de deleciones que interrumpen la pauta de lectura de la proteína, otros codones de *stop* o la falta de sitios de *splicing* fueron observadas. Por lo tanto, es posible que esta copia haya adquirido recientemente el codon de *stop* observado y se encuentre en proceso de pseudogenización, o que hubiera un error en la secuencia nucleotídica en esta región que produjera un cambio de aminoácido.

Las estima de la divergencia silenciosa y sinónima (tabla II.12) entre las duplicaciones detectadas en *D. ananassae* indican que la copia *ana2* es la más antigua debido a los valores más elevados de K_{sil} y K_s que presentan las comparaciones que incluyen esta copia.

El alineamiento de las tres copias del gen *Dca* (figura II.12A) muestra que la mayoría de los cambios aminoacídicos se encuentran en la zona central de la proteína mientras que las regiones N-terminal y C-terminal se mantienen más conservadas. La proteína codificada por el gen *Dca* presenta un total de 91, 71 y 62 reemplazamientos aminoacídicos en las comparaciones *ana1-ana2*, *ana1-ana3* y *ana2-ana3*, respectivamente. Algunos de estos reemplazamientos pueden deberse a la fijación de mutaciones no sinónimas ventajosas o a mutaciones selectivamente neutras. En el primer caso, los cambios en una de las copias podrían haber sido causados por evolución adaptativa debido a la aparición de una nueva función, mientras que en el segundo caso, la relajación de la selección purificadora en contra de las mutaciones no sinónimas en una de las copias habría permitido que la deriva llevara estas mutaciones a la fijación. Sin embargo, los valores observados de la relación K_a/K_s menores de 1 en todas las comparaciones apuntan a la relajación de la selección purificadora como la

causa principal de la fijación de los reemplazamientos aminoacídicos. No obstante, a pesar de la similitud observada en la distribución de la relación K_a/K_s a lo largo de la región codificadora en las tres posibles comparaciones, los menores valores de K_a/K_s en el exón 2 (figura II.13) sugieren una mayor restricción funcional en comparación al exón 1. Esto indica que los niveles de relajación de la selección purificadora son diferentes a lo largo del gen *Dca* en *D. ananassae*.

A diferencia de *D. ananassae*, las dos copias detectadas en *D. pseudoobscura* presentan unos niveles de divergencia silenciosa y no sinónima muy reducidos (tabla II.12) lo que sugiere un evento de duplicación relativamente reciente. De hecho, únicamente se detectaron dos reemplazamientos aminoacídicos fijados en las posiciones 117 y 140 del alineamiento (figura II.12B) entre las dos copias. Los niveles de la relación entre la divergencia no sinónima y sinónima ($K_a/K_s = 0.097$) sugieren que el gen *Dca* en *D. pseudoobscura* está bajo unos niveles de restricción funcional bastante elevados. Por otro lado, los resultados obtenidos a partir de los análisis de BLAST en *D. persimilis* detectaron una única copia del gen *Dca*, lo que sugiere que la duplicación se produjo después de la divergencia entre ambas especies. El gen de *D. persimilis* se encuentra ubicado en la misma región sinténica del gen *ps1* lo que sugiere que esta copia es la ancestral de las dos observadas en *D. pseudoobscura*.

CAPÍTULO III
EL GEN REGUCALCIN (RC)

III.1. POLIMORFISMO NUCLEOTÍDICO DEL GEN *RC*

III.1.1 RESULTADOS

III.1.1.1 POLIMORFISMO NUCLEOTÍDICO DEL GEN *RC* EN *D. MELANOGASTER*

Se secuenció un fragmento de aproximadamente 2.8 kb en 13 líneas de *D. melanogaster*. Dicho fragmento incluye toda la región codificadora del gen *Regucalcin* (*RC*) que comprende alrededor de 0.9 kb y parte de las regiones 5' y 3' flanqueantes con una longitud aproximada de 1 kb y 0.8 kb, respectivamente.

Análisis del polimorfismo nucleotídico

El alineamiento múltiple de las 13 líneas de *D. melanogaster* presenta un total de 2822 posiciones. Sin embargo, el número de posiciones se reduce a 2785 al eliminar aquellas posiciones con *gaps* en alguna o varias de las líneas analizadas. En este alineamiento se detectaron 8 polimorfismos por inserción/delección y 59 posiciones polimórficas por cambio nucleotídico (figura III.1).

En la región 5' flanqueante se detectaron 6 de los 8 polimorfismos por inserción/delección y 31 polimorfismos por cambio nucleotídico de los cuales 12 son únicos. En la región codificadora no se detectó ningún polimorfismo por inserción/delección. En esta región se identificaron 8 polimorfismos por cambio nucleotídico, todos ellos sinónimos. Tres de estos polimorfismos presentan variantes únicas. La región 3' flanqueante contiene los restantes 2 polimorfismos por inserción/delección y un total de 20 polimorfismos por cambio nucleotídico, 6 de ellos únicos.

Diversidad haplotípica

Al considerar los polimorfismos por cambio nucleotídico en toda la región analizada las 13 líneas de *D. melanogaster* se agrupan en 10 haplotipos. El número de haplotipos identificados no se ve afectado si se consideran asimismo los polimorfismos por

inserción/delección. De los 10 haplotipos detectados 9 corresponden a líneas únicas y uno está presente en cuatro líneas. La diversidad haplotípica es $H_d = 0.923$. Al considerar solamente los polimorfismos detectados en la región codificadora, las 13 líneas analizadas de *D. melanogaster* se agrupan en 3 haplotipos que representan una única variante proteica. Al tener en cuenta solamente los 31 polimorfismos detectados en la región 5' flanqueante, las 13 líneas de *D. melanogaster* se agrupan en un total de 6 haplotipos. En esta región la diversidad haplotípica es $H_d = 0.718$.

Desequilibrio de ligamiento y recombinación

El grado de desequilibrio de ligamiento se analizó para las posiciones polimórficas informativas detectadas. Entre los 38 posiciones polimórficas observadas, se realizaron 703 comparaciones de las cuales 261 (37.1%) son significativas por el test χ^2 . Ninguna de las comparaciones fue significativa al aplicar la corrección de Bonferroni. La región *RC* presenta valores significativos de los estadísticos Z_A ($Z_A = 0.786$; $P = 0.003$), Wall's B ($B = 0.6$; $P = 0.026$) y Wall's Q ($Q = 0.789$; $P = 0.001$) bajo la premisa conservativa de no recombinación. Se detectaron un mínimo de 5 eventos de recombinación (Rm): 2 en la región 5' flanqueante, 1 entre la región 5' flanqueante y la región codificadora y 2 en la región 3' flanqueante.

En la región 5' flanqueante, de las 171 comparaciones realizadas 155 (90.6%) son significativas por el test χ^2 y ninguna tras aplicar la corrección de Bonferroni. Para la región codificadora, las 10 comparaciones (100%) realizadas son significativas con la corrección de Bonferroni. En la región 3' flanqueante el 72.5% de las comparaciones son significativas por el test χ^2 y el 42.8% con la corrección de Bonferroni. Los estadísticos Z_A , Wall's B y Q son significativos en la región 5' flanqueante con valores de $Z_A = 0.822$ ($P = 0.003$), Wall's $B = 0.611$ ($P = 0.031$) y Wall's $Q = 0.789$ ($P = 0.002$). En la región codificadora únicamente el estadístico Z_A es significativo ($Z_A = 0.803$; $P = 0.043$) mientras que los valores de los tres estadísticos son significativos en la 3' flanqueante ($Z_A = 0.827$; $P = 0.012$; Wall's $B = 0.692$; $P = 0.013$; Wall's $Q = 0.785$; $P = 0.016$).

Diversidad nucleotídica

Los niveles del polimorfismo nucleotídico fueron estimados para distintos tipos de posiciones (tabla III.1). La diversidad nucleotídica π para toda la región analizada es 0.007. Esta estima es similar a la de 0.009 obtenida para las posiciones silenciosas (regiones 5' y 3' flanqueantes, intrón y posiciones sinónimas de la región codificadora). Dentro de la región codificadora, la diversidad nucleotídica para las posiciones sinónimas (π_s) es 0.009 mientras que para las posiciones no sinónimas (π_a) es 0. En la región no codificadora

(regiones 5' y 3' flanqueantes e intrón) esta estima es 0.009. Las estimas de la diversidad nucleotídica fueron muy similares a las estimas de θ para todos los tipos de posiciones.

Tabla III.1. Estimaciones del polimorfismo nucleotídico del gen *RC* en *D. melanogaster*

Región	Nº de Sitios	S	π	θ
No Codificadora	1876	51	0.009	0.009
5'	979	31	0.009	0.010
3'	840	20	0.009	0.007
Codificadora	909	8	0.002	0.003
Silenciosa	2102.67	59	0.009	0.009
Sinónima	226.67	8	0.009	0.011
No sinónima	682.33	0	0	0
Total	2785	59	0.007	0.007

Test del neutralismo

Los test propuestos por Tajima (1989) y Fu y Li (1993) se aplicaron para determinar si los patrones del polimorfismo observados en la región estudiada se ajustaban al neutralismo (tabla III.2). Estos tests se realizaron independientemente para diferentes regiones y tipos de posiciones. En la región 5' flanqueante todos los estadísticos propuestos para aplicar los distintos tests fueron negativos pero ninguno de ellos fue significativo. Tampoco fueron significativos los tests realizados en el total de la región no codificadora o para las posiciones silenciosas. Los resultados obtenidos con los tests de Tajima y Fu y Li indican que el patrón del polimorfismo en la región estudiada se ajusta a las predicciones del neutralismo.

El test H de Fay y Wu (2000) fue utilizado para determinar si existe un exceso de mutaciones derivadas segregando a alta frecuencia en la región analizada (tabla III.2). El estadístico H se estimó para diferentes tipos de posiciones independientemente. Para poder determinar el estado de los cambios observados se utilizó la secuencia de *D. simulans* como especie *outgroup*. Este test es negativo y estadísticamente significativo ($P < 0.05$) en todos los casos excepto en la región 5' flanqueante donde es marginalmente significativo ($0.1 > P > 0.05$). Por lo tanto, este resultado indica que existe un exceso de mutaciones derivadas segregando a alta frecuencia en toda la región analizada.

Tabla III.2. Tests del neutralismo en el gen *RC* de *D. melanogaster*

Test	No codificadora	5'	Silenciosas	Total
Tajima <i>D</i>	0.175	- 0.424	0.057	0.058
<i>P(D < Dobs)</i>	0.615	0.365	0.565	0.563
Fu y li <i>D</i>*	- 0.018	- 0.174	- 0.032	- 0.032
<i>P(D* < D*obs)</i>	0.463	0.419	0.465	0.456
Fu y Li <i>F</i>*	0.038	- 0.276	- 0.009	0.009
<i>P(F* < F*obs)</i>	0.494	0.365	0.467	0.473
Fay y Wu <i>H</i>	- 13.192	- 8.333	- 16.333	- 16.333
<i>P(H < Hobs)</i>	0.034*	0.053*	0.03*	0.037*

Probabilidad de obtener un valor del test estadístico menor que el observado tras simulación de coalescencia sin recombinación.

* $P < 0.05$

$0.1 > P > 0.05$

El test de McDonald y Kreitman (1991) se aplicó para contrastar si la relación entre el número de polimorfismos dentro de especie y el número de diferencias fijadas entre especies es equivalente para los cambios sinónimos y los no sinónimos. Para la comparación interespecífica se utilizó *D. simulans*. No se detectaron diferencias significativas entre ambas relaciones (tabla III.3).

Tabla III.3. Test de MK (1991) en el gen *RC* de *D. melanogaster*

	Polimorfismo en <i>D. melanogaster</i>	Diferencias fijadas con <i>D. simulans</i>
No Sinónimo	0	4
Sinónimo	8	22
		$\chi^2 = 0.306$
		$P = 0.58$

Genealogía de los alelos

La genealogía de los diferentes alelos analizados se muestra en la figura III.2. Para este análisis se utilizó *D. simulans* como especie *outgroup*. En el árbol se observan los

cuatro alelos que comparten el mismo haplotipo (CN8, CN26, CN28 y CN39) así como el resto de haplotipos presentes en una sola línea.

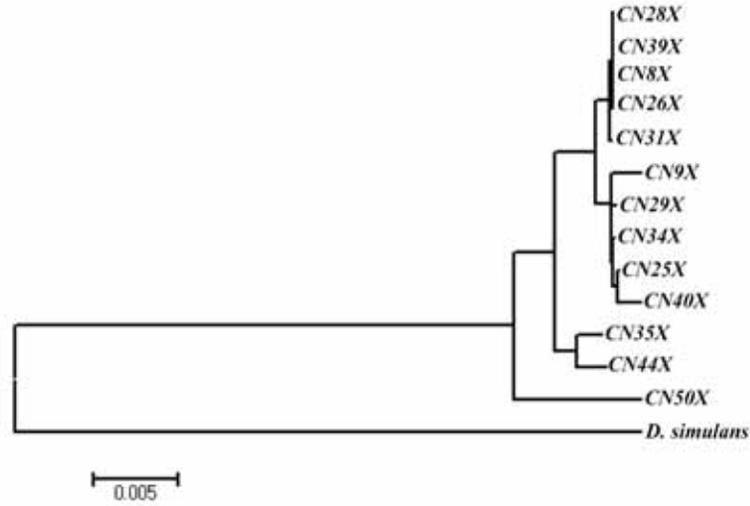


Figura III.2. Genealogía de los alelos de la población de Sant Sadurní d'Anoia (CN) reconstruida por el método de *neighbor-joining* (Saitou y Nei, 1987) usando la distancia de Kimura de 2 parámetros (Kimura 1980).

III.1.1.2 POLIMORFISMO NUCLEOTÍDICO DEL GEN *RC* EN *D. SUBOBSCURA*

Se secuenció una región de aproximadamente 1.5 kb en 12 líneas de *D. subobscura* de una población natural de Riba-roja d'Ebre (Catalunya). El fragmento secuenciado incluye toda la región codificadora del gen *Regucalcin* (*RC*) así como parte de las regiones 5' y 3' flanqueantes con una longitud aproximada de 10 pb y 500 pb, respectivamente.

Ubicación citológica del gen *RC*

Los resultados de la hibridación *in situ* del gen *RC* en los cromosomas politénicos de *D. subobscura* muestran una única señal ubicada en la banda citológica 12A del cromosoma A (figura III.3).

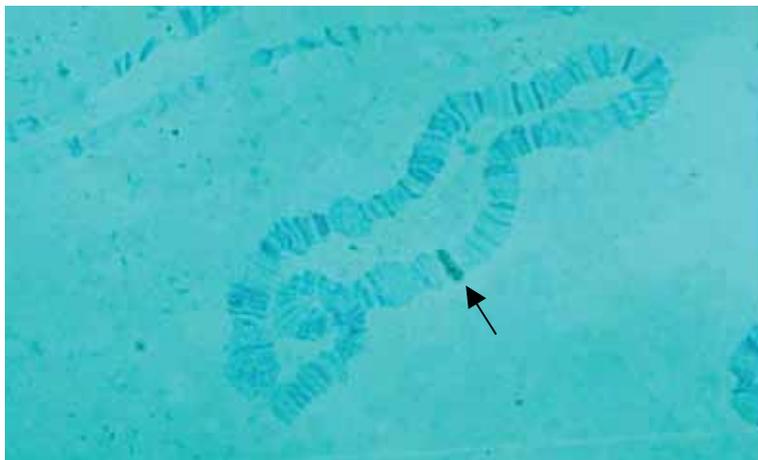


Figura III.3. Hibridación *in situ* del gen *RC* en los cromosomas politénicos de *D. subobscura*. La flecha indica la señal de hibridación.

Análisis del polimorfismo nucleotídico

El alineamiento múltiple de las 12 líneas de *D. subobscura* comprende un total de 1523 posiciones. Este número se reduce a 1521 al eliminar las posiciones que presentan *gaps* en una o varias de las líneas analizadas. En este alineamiento se detectaron tanto polimorfismos por cambio nucleotídico como polimorfismos por inserción/delección de

nucleótidos. Un total de 19 polimorfismos por cambio nucleotídico y 2 por inserción/delección fueron identificadas en toda la región analizada (figura III.4).

En la región 5' flanqueante no se identificó ningún polimorfismo ni por cambio nucleotídico ni por inserción/delección. En la región codificadora se detectaron 6 posiciones polimórficas todas ellas con cambios nucleotídicos sinónimos y de las cuales 4 presentan variantes únicas. En la región 3' flanqueante se identificaron los restantes 13 polimorfismos por cambio nucleotídico (7 de ellos con variantes únicas) y los 2 por inserción/delección de nucleótidos (figura III.4).

	167	236	596	650	769	770	772	880	964	1016	1068	1081	1088	1090	1111	1125	1215	1229	1463	1465	1470
Sub230	C	T	T	G	T	C	C	T	G	G	T	A	C	A	C	G	C	A	G	C	C
Sub232	C	.	.
Sub303	T	T	T
Sub305	.	C	C	A	.	.	G	.	A	.	.	.	A	.	A	T
Sub306	.	.	C	.	A	T	A
Sub309	.	.	C	.	A	T	A
Sub314	G	.	A	.	.	.	A	.	A
Sub317	.	.	C	.	.	.	G	G	.	.	.
Sub337	.	.	C	C	A	.	.
Sub342	.	.	C
Sub345	.	.	C	.	A	T	A
Sub354	G	A	A	C	.	.
<i>D. madeirensis</i>	.	.	C	A	?	?	?
<i>D. guanche</i>	.	.	C	?	?	?	?	?	?	?	?	?	?	?	?	?

Figura III.4. Polimorfismo nucleotídico en el gen *Regucalcin* en las 12 líneas secuenciadas de *D. subobscura*. Las posiciones polimórficas en la región codificadora se indican sombreadas en gris. Los puntos (.) indican nucleótidos idénticos a los de la secuencia de referencia, los guiones (-) deleciones de nucleótidos respecto a esta secuencia y el signo de interrogación (?) indica posiciones no analizadas. Las dos últimas filas muestran la información en *D. guanche* y *D. madeirensis* para las posiciones identificadas como polimórficas en *D. subobscura*.

Desequilibrio de ligamiento y recombinación

El grado de desequilibrio de ligamiento de la región estudiada se analizó entre las posiciones polimórficas informativas detectadas. De las 28 comparaciones entre los 8 polimorfismos detectados, 8 (28.57%) son significativas por el test de χ^2 y 2 (7.14%) tras

aplicar la corrección de Bonferroni. Se detectaron 2 eventos de recombinación ($R_m = 2$) entre las posiciones 596-772 y 772-964.

Diversidad nucleotídica

Los niveles del polimorfismo nucleotídico fueron estimados para distintos tipos de posiciones (tabla III.4). La diversidad nucleotídica silenciosa (π_{sil}) estimada fue de 0.006, ligeramente inferior que la detectada en *D. melanogaster* ($\pi_{sil} = 0.009$, tabla III.1). Dentro de la región codificadora, la estima para las posiciones sinónimas (π_s) es 0.007 mientras que para las posiciones no sinónimas (π_a) es 0. En la región no codificadora (regiones 5' y 3' flanqueantes e intrón) la estima de 0.005 es inferior a la diversidad nucleotídica sinónima ($\pi_s = 0.007$). Las estimas de θ para la mayoría de posiciones son más altas que las estimas de la diversidad nucleotídica. Este resultado indicaría un exceso de polimorfismos con variantes segregando a bajas frecuencias en la región analizada.

Tabla III.4. Estimaciones del polimorfismo nucleotídico del gen *RC* en *D. subobscura*

Región	Nº de Sitios	S	π	θ
No Codificadora	612	13	0.005	0.007
Codificadora	909	6	0.002	0.002
Silenciosa	612	19	0.006	0.007
Sinónima	221	6	0.007	0.009
No sinónima	688	0	0	0
Total	1521	19	0.003	0.004

Test del neutralismo

Los test propuestos por Tajima (1989) y Fu y Li (1993) se emplearon para detectar posibles desviaciones del modelo neutro en el patrón de polimorfismo observado en el gen *RC* (tabla III.5). Estos tests fueron realizados independientemente para las posiciones no codificadoras y para las posiciones silenciosas. Los estadísticos propuestos para aplicar los distintos tests son negativos en todos los casos, sin embargo ninguno de ellos presenta

valores significativos. De esta forma, los resultados obtenidos muestran que el patrón de polimorfismo en la región estudiada se ajusta a las predicciones del neutralismo.

El test H de Fay y Wu (2000) se aplicó para determinar si existe un exceso de mutaciones derivadas segregando a elevada frecuencia en la región analizada (tabla III.5). El estadístico H se estimó para toda la región analizada y para las posiciones no codificadoras y silenciosas independientemente. En el análisis se utilizó *D. madeirensis* como especie *outgroup*. A diferencia de *D. melanogaster*, este test es en todos los casos positivo y no significativo. Este resultado señala que no existe un exceso de mutaciones derivadas segregando a alta frecuencia en la región analizada.

Tabla III.5. Tests del neutralismo en el gen *RC* de *D. subobscura*

Test	No codificadora	Silenciosas	Total
Tajima <i>D</i>	- 0.985	- 0.961	- 0.961
<i>P(D < D_{obs})</i>	0.157	0.168	0.178
Fu y Li <i>D</i>*	- 0.728	- 0.930	- 0.930
<i>P(D* < D*_{obs})</i>	0.294	0.203	0.202
Fu y Li <i>F</i>*	- 0.904	- 1.068	- 1.068
<i>P(F* < F*_{obs})</i>	0.231	0.190	0.204
Fay y Wu <i>H</i>	0.727	1.757	1.757
<i>P(H < H_{obs})</i>	0.466	0.632	0.615

Probabilidad de obtener un valor del test estadístico menor que el observado tras simulación de coalescencia sin recombinación.

El test de MacDonal y Kreitman (1991) se aplicó para detectar si la relación entre el número de diferencias no sinónimas y sinónimas fijadas entre especies es equivalente a la relación entre el número de polimorfismos no sinónimos y sinónimos detectados dentro de especie. Para realizar la comparación interespecífica se utilizaron *D. guanche*, *D. madeirensis* y *D. pseudoobscura* independientemente. Para las dos primeras comparaciones se detectaron diferencias significativas entre ambas relaciones (tabla III.6). La desviación observada corresponde a un exceso de cambios no sinónimos fijados en los linajes de *D. guanche* y *D. madeirensis*. De hecho en este último linaje la relación d_n/d_s es mayor que 1.

Tabla III.6. Test de MK (1991) en el gen *RC* de *D. subobscura*

	Polimorfismo en		Diferencias fijadas	
	<i>D. subobscura</i>	<i>D. guanche</i>	<i>D. madeirensis</i>	<i>D. pseudoobscura</i>
No Sinónimo	0	12	12	33
Sinónimo	6	15	5	71
		$\chi^2 = 4.190$ $P = 0.0406^*$	$\chi^2 = 8.856$ $P = 0.002^{**}$	$\chi^2 = 1.419$ $P = 0.233$

El test de HKA (Hudson et al. 1987) se aplicó para detectar una posible desviación respecto a lo esperado de acuerdo al neutralismo en la relación entre el nivel de polimorfismo silencioso y el nivel de divergencia silenciosa entre *D. subobscura* y *D. guanche* en la región *RC*. Para aplicar este test, los niveles de polimorfismo y divergencia silenciosa del gen *RC* se compararon con los niveles de la región *rp49* de una población de Galicia con la ordenación O_{3+4} (Rozas et al. 1995) asumiendo que esta región evoluciona de acuerdo al neutralismo. El valor obtenido del estadístico χ^2 en el test HKA no es significativo ($\chi^2 = 1.649$, $P = 0.199$).

Genealogía de los alelos

La genealogía obtenida para los diferentes alelos de *D. subobscura* se muestra en la figura III.5. Para este análisis se utilizó la secuencia de *D. madeirensis* como especie *outgroup*. En el árbol se identifican los dos haplotipos compartidos por dos líneas (Sub232-Sub230 y Sub306-Sub309) así como los 8 haplotipos restantes presentes en líneas individuales.

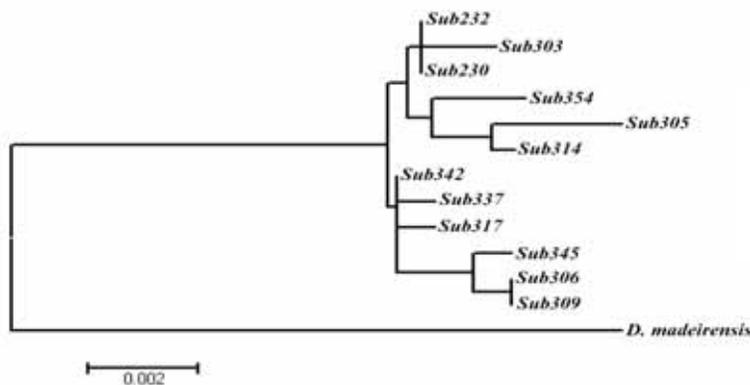


Figura III.5. Genealogía de los alelos de *D. subobscura* reconstruida por el método de *neighbor-joining* (Saitou y Nei, 1987) usando la distancia de Kimura de 2 parámetros (Kimura 1980).

III.1.2 DISCUSIÓN

III.1.2.1 DIVERSIDAD NUCLEOTÍDICA DEL GEN *RC* EN *D. MELANOGASTER* Y *D. SUBOBSCURA*

Los niveles de diversidad nucleotídica silenciosa y sinónima ($\pi_{\text{sil}} = \pi_{\text{s}} = 0.009$) observados en el gen *RC* de *D. melanogaster* son muy similares a los detectados en *D. subobscura* ($\pi_{\text{sil}} = 0.006$ y $\pi_{\text{s}} = 0.007$). Además, en ninguna de las dos especies se detectaron polimorfismos no sinónimos, resultado que sugiere que el gen *RC* se encuentra bajo la acción de una fuerte selección purificadora en contra de las mutaciones que provocan cambios aminoacídicos en la proteína codificada por el gen.

Por otro lado, los niveles de variación del gen *RC* en *D. melanogaster* y *D. subobscura* pueden considerarse típicos de ambas especies. Así, las estimas de la diversidad nucleotídica silenciosa y sinónima ($\pi_{\text{sil}} = \pi_{\text{s}} = 0.009$) del gen *RC* en *D. melanogaster* son similares a las estimas promedio de la diversidad nucleotídica silenciosa ($\pi_{\text{sil}} = 0.010$) en 9 genes ligados al cromosoma X en *D. melanogaster* (Moriyama y Powell 1996) y a la estima promedio de la diversidad sinónima ($\pi_{\text{s}} = 0.011$) de 5 genes ubicados en el cromosoma X de poblaciones no africanas de *D. melanogaster* (Andolfatto 2001). Finalmente, la diversidad nucleotídica en las regiones no codificadoras ($\pi_{\text{NC}} = 0.005$) del gen *RC* también es similar a la diversidad nucleotídica promedio ($\pi_{\text{NC}} = 0.0038$) obtenida en el análisis multilocus realizado en la misma población (Sant Sadurní d'Anoia) de *D. melanogaster* y que incluye 109 regiones no codificadoras del cromosoma X (Orengo y Aguadé 2004). En el caso de *D. subobscura*, las estimas de la diversidad nucleotídica silenciosa del gen *RC* ($\pi_{\text{sil}} = 0.006$) también son similares a las estimas promedio encontradas en genes ubicados en el cromosoma X analizados en esta especie ($\pi_{\text{sil}} = 0.007$, ver tabla I.8).

No obstante, la variabilidad nucleotídica del gen *RC* muestra características diferenciales entre *D. melanogaster* y *D. subobscura*. En *D. melanogaster* y a diferencia de lo que ocurre en *D. subobscura*, la región *RC* presenta un porcentaje considerablemente elevado de asociaciones significativas entre los polimorfismos informativos detectados (37.1% de comparaciones significativas por el test χ^2), así como una fuerte estructura de la variación nucleotídica tal como indican los valores significativos que presentan los

estadísticos Z_A , *Wall's B* y Q incluso bajo la premisa conservativa de no recombinación. Los altos niveles de desequilibrio de ligamiento y de estructuración de la variación genética pueden ser causados por los procesos demográficos sufridos durante la expansión de *D. melanogaster* o por los eventos selectivos que han sido importantes en la adaptación de esta especie a nuevos ambientes (Andolfatto y Przeworski 2000; Pritchard y Przeworski 2001; Gilad et al. 2002).

Otra característica a destacar de la variación nucleotídica del gen *RC* en *D. melanogaster* es el exceso significativo de mutaciones derivadas segregando a elevadas frecuencias observado al aplicar el test H de Fay y Wu (2000). Este resultado podría indicar un barrido selectivo en o cerca de la región *RC*. Tras un barrido selectivo las mutaciones derivadas que se encuentran ligadas al alelo seleccionado pueden llegar a alcanzar elevadas frecuencias. De hecho, si el arrastre selectivo es incompleto o por causa de la recombinación muchas de las mutaciones derivadas no llegan a fijarse en la población y permanecen segregando a elevadas frecuencias. No obstante, los barridos selectivos también producen valores negativos significativos de los estadísticos propuestos para aplicar los tests de Tajima y de Fu y Li, resultado que no se ha detectado en la región *RC*. Por otro lado, estudios recientes empleando simulaciones de coalescencia de modelos demográficos (Przeworski 2001; Wakeley y Aliacar 2001) sugieren que el exceso de mutaciones derivadas no es una característica exclusiva de la acción de la selección positiva puesto que cuellos de botella recientes o modelos de metapoblaciones también pueden producir frecuencias altas de mutaciones derivadas en comparación a las esperadas bajo el modelo neutro. Sin embargo, tal y como se ha comentado anteriormente, el análisis por simulaciones de eventos demográficos realizado por Orengo y Aguadé (2004) en la población de Sant Sadurní d'Anoia parecía indicar que el nivel y patrón de polimorfismo nucleotídico en esta población sólo era compatible con una baja probabilidad con un cuello de botella de intensidad intermedia relativamente reciente. Por lo tanto, parece poco probable que, por lo menos, los efectos demográficos debidos a cuellos de botella poblacionales puedan explicar el exceso de mutaciones derivadas segregando a elevada frecuencia detectado en la región *RC* de *D. melanogaster*.

La característica más destacable del estudio del gen *RC* en *D. subobscura* es el resultado significativo que se obtiene al aplicar el test de McDonald y Kreitman (1991). Este resultado indica que la relación entre el número de diferencias no sinónimas y sinónimas fijadas entre especies no es equivalente a la relación entre el número de polimorfismos no sinónimos y sinónimos detectados dentro de especie. De hecho, existen más diferencias no sinónimas fijadas entre *D. subobscura* y *D. guanche* o *D. madeirensis* que las esperadas. Este exceso de sustituciones de reemplazamiento puede ser el resultado de la fijación de mutaciones selectivamente ventajosas en los linajes que conducen a *D. guanche* y *D. madeirensis*. En la comparación entre *D. subobscura* y *D. guanche* las 12 sustituciones no sinónimas detectadas se encuentran en el primer exón. Al comparar *D. subobscura* y *D. madeirensis*, 9 de las 12 sustituciones también se hallan en el primer exón y las 3 restantes en el segundo. Esto sugiere que la evolución adaptativa ha causado una mayor diferenciación en la región N-terminal de la proteína que en la región C-terminal.

III.2 DIVERGENCIA DEL GEN *RC*

III.2.1 RESULTADOS

III.2.1.1 DIVERGENCIA NUCLEOTÍDICA DEL GEN *RC* EN *DROSOPHILA*

La región codificadora del gen *RC* fue secuenciada en 7 especies de *Drosophila*: cuatro pertenecientes al subgrupo *melanogaster* (*D. melanogaster*, *D. simulans*, *D. yakuba* y *D. erecta*) y 3 al grupo *obscura* (*D. subobscura*, *D. madeirensis* y *D. guanche*). Para realizar los diferentes análisis, se incluyeron además las secuencias nucleotídicas de la región codificadora del gen *RC* obtenidas a partir de la base de datos de las especies *D. sechellia*, *D. ananassae* y *D. pseudoobscura*.

El alineamiento múltiple de la secuencia nucleotídica de la región codificadora de las 10 especies comprende un total de 909 pb. No se detectaron deleciones o inserciones en ninguna de las secuencias. La estructura del gen *RC*, al igual que la observada en el gen *Dca*, está formada por dos exones separados por un pequeño intrón. La longitud de la región codificadora incluyendo el intrón varía desde 966 pb en *D. melanogaster* hasta 992 pb en *D. subobscura*.

En la tabla III.7 se presentan las estimas de la divergencia de acuerdo al método de Nei-Gojobori (1986) modificado por Nei y Kumar (2000) para las posiciones sinónimas (K_s) arriba de la diagonal y para las posiciones no sinónimas (K_a) debajo de la diagonal. En esta tabla se observa que la divergencia sinónima es superior a la divergencia no sinónima en todas las comparaciones. Las estimas de la divergencia para las posiciones sinónimas entre las especies del grupo *melanogaster* y del grupo *obscura* son similares en todas las comparaciones y oscilan entre 0.529 y 0.660. Los niveles de *codon bias* para cada una de las especies se muestran en la tabla III.8. Las especies del grupo *melanogaster* presentan el mayor sesgo en el uso de codones sinónimos, con valores de ENC desde 31.004 en *D. ananassae* hasta 35.265 en *D. melanogaster*. En las especies restantes los valores detectados oscilan entre 38.822 en *D. pseudoobscura* y 41.938 en *D. guanche*.

Tabla III.7. Divergencia sinónima (K_s) y no sinónima (K_a) en el gen *RC*

	<i>mel</i>	<i>sim</i>	<i>sech</i>	<i>yak</i>	<i>ere</i>	<i>ana</i>	<i>pseu</i>	<i>sub</i>	<i>made</i>	<i>guan</i>
<i>mel</i>		0.098	0.115	0.170	0.193	0.452	0.597	0.658	0.659	0.660
<i>sim</i>	0.006		0.037	0.151	0.151	0.390	0.573	0.577	0.578	0.579
<i>sech</i>	0.014	0.011		0.165	0.160	0.414	0.576	0.563	0.564	0.557
<i>yak</i>	0.010	0.017	0.025		0.085	0.360	0.561	0.541	0.542	0.536
<i>ere</i>	0.006	0.009	0.017	0.011		0.353	0.547	0.559	0.559	0.529
<i>ana</i>	0.052	0.056	0.062	0.058	0.056		0.570	0.533	0.541	0.531
<i>pseu</i>	0.060	0.058	0.066	0.066	0.059	0.084		0.326	0.332	0.322
<i>sub</i>	0.104	0.110	0.117	0.114	0.108	0.116	0.056		0.023	0.062
<i>made</i>	0.100	0.106	0.113	0.111	0.105	0.114	0.052	0.019		0.075
<i>guan</i>	0.103	0.109	0.118	0.114	0.110	0.116	0.055	0.019	0.024	

Divergencia por nucleótido corregida por el método de Nei-Gojobori (1986) modificado por Nei y Kumar (2000). Estimaciones de la divergencia sinónima (K_s) y no sinónima (K_a) se muestran arriba y bajo la diagonal, respectivamente. *mel*, *D. melanogaster*; *sim*, *D. simulans*; *sech*, *D. sechellia*; *mau*, *D. mauritiana*; *yak*, *D. yakuba*; *ere*, *D. erecta*; *ana*, *D. ananassae*; *pse*, *D. pseudoobscura*; *sub*, *D. subobscura*; *mad*, *D. madeirensis* y *gua*, *D. guanche*.

Tabla III.8. Estimaciones de *codon bias* en el gen *RC*

Especies	Scaled χ^2	CBI	ENC
Grupo <i>melanogaster</i>			
<i>D. melanogaster</i>	0.806	0.670	35.265
<i>D. simulans</i>	0.980	0.712	32.117
<i>D. sechellia</i>	0.977	0.684	32.284
<i>D. yakuba</i>	0.917	0.664	33.177
<i>D. erecta</i>	1.023	0.690	32.049
<i>D. ananassae</i>	0.989	0.768	31.004
Grupo <i>obscura</i>			
<i>D. pseudoobscura</i>	0.539	0.577	38.822
<i>D. subobscura</i>	0.462	0.516	41.599
<i>D. madeirensis</i>	0.475	0.522	40.226
<i>D. guanche</i>	0.460	0.528	41.938

CBI, índice de *codon bias*; ENC, número efectivo de codones

A diferencia de lo observado al analizar la divergencia sinónima, las estimaciones de K_a entre las especies del grupo *melanogaster* y las especies paleárticas del grupo *obscura* (que oscilan entre 0.10 y 0.118) son superiores a las detectadas entre las especies del grupo *melanogaster* y *D. pseudoobscura* (desde 0.058 hasta 0.084). La figura III.6A muestra el árbol obtenido con el método de *neighbor-joining* utilizando K_a como distancia genética. En este árbol destaca la longitud de la rama que conduce a los linajes de las especies *D. subobscura*, *D. madeirensis* y *D. guanche*. Además la

agrupación de estas tres especies en este árbol no coincide con la obtenida en el árbol reconstruido en base a la divergencia sinónima, que de hecho reproduce las relaciones filogenéticas aceptadas para las especies del *cluster subobscura*.

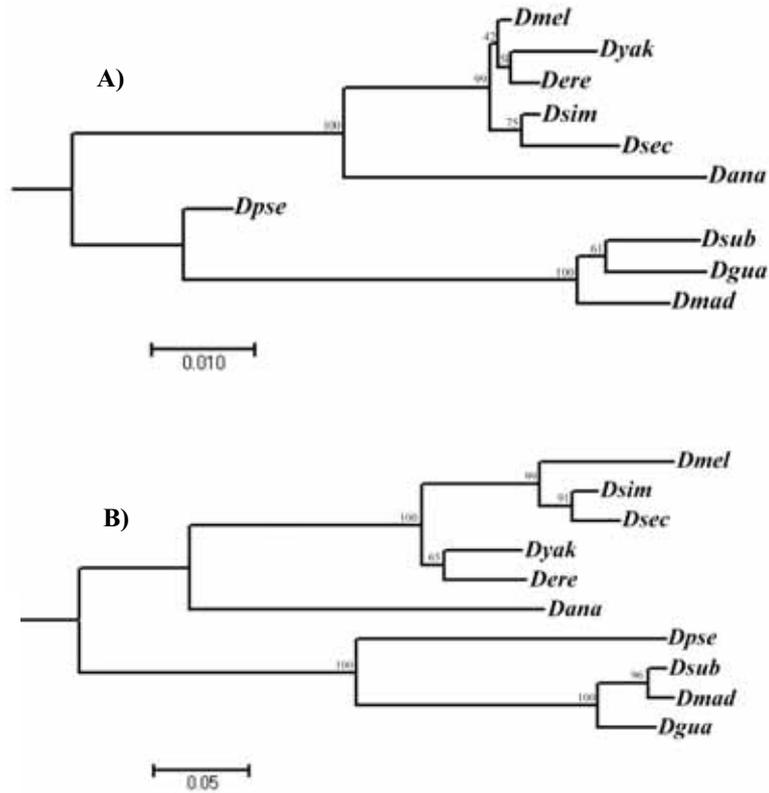


Figura III.6. Árboles filogenéticos obtenidos por el método de *neighbor-joining* (Saitou y Nei 1987) en base a A) la divergencia no sinónima y B) la divergencia sinónima.

Para contrastar la constancia en la tasa de sustitución nucleotídica en los linajes que conducen a las especies del grupo *obscura* se aplicaron *relative-rate tests*. En la tabla III.9 se presentan los resultados obtenidos al aplicar el *relative-rate test* propuesto por Wu y Li (1985) para todas las sustituciones nucleotídicas detectadas en la región codificadora del gen *RC*. Los análisis se realizaron empleando las especies del grupo *melanogaster* como *outgroup*. En las comparaciones usando *D. melanogaster*, *D. simulans*, *D. sechelia*, *D. yakuba* y *D. erecta*, se detectó una diferencia altamente significativa ($P < 0.005$) en la tasa de sustitución en el linaje de *D. pseudoobscura* en comparación con cualquiera de las restantes especies del grupo *obscura*. Sin embargo, las comparaciones utilizando *D. ananassae* como especie *outgroup* no son

significativas, aunque muestran también una aceleración en el tasa de sustitución en los linajes que conducen a las especies paleárticas del grupo *obscura*.

Tabla III.9. Relative-Rate Test

Out-Group	Especie 1- Especie 2	Test de Wu y Li			Test de Tajima					
		$K_{12} - K_{23}$	SD	z - score	Sustituciones sinónimas			Sustituciones no sinónimas		
					m1	m2	χ^2	r1	r2	χ^2
mel	<i>pseu-sub</i>	-0.046695	0.014557	-3.207722**	28	34	0.581	3	29	21.125***
	<i>pseu-made</i>	-0.043663	0.014459	-3.019875**	29	35	0.562	3	27	19.2***
	<i>pseu-guan</i>	-0.046695	0.014483	-3.224225**	25	32	0.859	4	29	18.939***
sim	<i>pseu-sub</i>	-0.042850	0.014390	-2.977740**	30	31	0.016	1	30	27.129***
	<i>pseu-made</i>	-0.039863	0.014294	-2.788738**	31	32	0.015	1	28	25.137***
	<i>pseu-guan</i>	-0.042850	0.014317	-2.993008**	28	29	0.017	1	30	27.129***
sech	<i>pseu-sub</i>	-0.038960	0.014417	-2.702332**	31	29	0.066	2	31	25.484***
	<i>pseu-made</i>	-0.035963	0.014321	-2.511271*	32	31	0.015	2	28	22.533***
	<i>pseu-guan</i>	-0.038960	0.014344	-2.716134**	29	28	0.017	2	30	24.5***
yak	<i>pseu-sub</i>	-0.035742	0.014357	-2.489460*	32	30	0.064	2	29	23.516***
	<i>pseu-made</i>	-0.032769	0.014263	-2.297539*	33	31	0.062	2	27	21.551***
	<i>pseu-guan</i>	-0.034382	0.014281	-2.407565*	31	27	0.275	2	30	24.5***
ere	<i>pseu-sub</i>	-0.042782	0.014325	-2.986617**	31	32	0.015	1	30	27.129***
	<i>pseu-made</i>	-0.039811	0.014229	-2.797878**	32	33	0.015	1	28	25.137***
	<i>pseu-guan</i>	-0.038375	0.014207	-2.7001084**	31	28	0.152	1	31	28.125***
ana	<i>pseu-sub</i>	-0.019158	0.014464	-1.324567	33	31	0.062	6	21	8.333*
	<i>pseu-made</i>	-0.019256	0.014409	-1.336332	32	30	0.064	7	22	7.758*
	<i>pseu-gua</i>	-0.019158	0.014389	-1.331398	31	26	0.438	7	25	10.125*

m1 y m2, número de sustituciones sinónimas en los linajes que llevan a las especies 1 y 2, respectivamente.

r1 y r2, número de sustituciones no sinónimas en los linajes que llevan a las especies 1 y 2, respectivamente.

Nombre de las especies como en la tabla III.7.

* $0.05 > P > 0.01$

** $0.01 > P > 0.001$

*** $P < 0.001$

Con el propósito de identificar que clase de sustitución es la responsable de la desviación en la constancia de la tasa de sustitución nucleotídica, se aplicó el *relative rate test* de Tajima (1993) independientemente para las sustituciones sinónimas y no sinónimas (tabla III.9). En todas las comparaciones, incluso en las efectuadas con *D. ananassae* como *outgroup*, la desviación observada es causada por diferencias en la tasa de fijación de mutaciones no sinónimas. Por lo tanto, los resultados obtenidos a partir de los *relative-rate tests* indican que las especies *D. subobscura*, *D. madeirensis* y *D. guanche*, presentan un incremento significativo en la tasa de sustitución no sinónima en comparación a *D. pseudoobscura*.

Evolución del intrón

En todas las especies analizadas la estructura del gen *RC* está formada por dos exones separados por un pequeño intrón. De la misma forma que el gen *Dca*, este intrón

se encuentra ubicado entre la segunda y la tercera base de un codón por lo que puede ser clasificado de fase 2. Los análisis realizados para determinar una posible diferencia en los patrones de sustitución nucleotídica, de inserciones y de deleciones en el intrón se efectuaron independientemente en el *cluster melanogaster* y en el *cluster subobscura*.

Dentro del subgrupo *melanogaster*, el intrón más largo lo presentan las especies *D. simulans* y *D. sechelia* con 62 pb, mientras que dentro del grupo *obscura*, *D. subobscura* presenta el mayor intrón (83 pb). El promedio del tamaño para las especies del subgrupo *melanogaster* (60.2 pb) es menor que el observado para las especies incluidas en el grupo *obscura* (79.5 pb). Sin embargo, el intrón presente en todas las especies puede catalogarse dentro de la clase de intrones cortos (Mount *et al.* 1992; Yu *et al.* 2002). En la tabla III.10 se presentan el tamaño del intrón en cada especie y las estimas de la divergencia dentro del subgrupo *melanogaster* y el grupo *obscura*.

Tabla III.10. Tamaño del intrón en el gen *RC*

Especies	Tamaño	Divergencia intrón
Subgrupo <i>melanogaster</i>		
<i>mel-sim</i>	57-62 (62)	0.134
<i>mel-sech</i>	57-62 (62)	0.093
<i>mel-yak</i>	57-61 (62)	0.419
<i>mel-ere</i>	57-59 (62)	0.290
<i>sim-sech</i>	62-62 (62)	0.050
<i>sim-yak</i>	62-61 (62)	0.322
<i>sim-ere</i>	62-59 (62)	0.237
<i>sech-yak</i>	62-61 (62)	0.322
<i>sech-ere</i>	62-59 (62)	0.214
<i>yak-ere</i>	61-59 (62)	0.195
Grupo <i>obscura</i>		
<i>pseu-sub</i>	81-83 (84)	0.495
<i>pseu-made</i>	81-77 (82)	0.506
<i>pseu-guan</i>	81-77 (82)	0.533
<i>sub-made</i>	83-77 (83)	0
<i>sub-guan</i>	83-77 (83)	0.040
<i>made-guan</i>	77-77 (77)	0.040

Nombre de las especies como en la tabla III.7.

Entre parentesis (); longitud del alineamiento entre pares de especies.

El número de sustituciones nucleotídicas y de *indels* que presenta la secuencia del intrón en las especies del *cluster melanogaster* (*D. melanogaster*, *D. simulans* y *D. sechellia*) y las incluidas en el *cluster subobscura* (*D. subobscura*, *D. madeirensis* y *D. guanche*) se infirieron por el criterio de máxima parsimonia (tabla III.11). *D. yakuba* y

D. erecta fueron utilizadas como especies *outgroup* en la comparación dentro del *cluster melanogaster* y *D. pseudoobscura* fue utilizada en el caso del *cluster subobscura*.

En el *cluster melanogaster* se detectaron 2 *indels* y 18 sustituciones nucleotídicas al utilizar *D. yakuba* como *outgroup* y 2 *indels* y 13 sustituciones al utilizar *D. erecta*. El *cluster subobscura* presenta 3 *indels* y 30 sustituciones nucleotídicas. A diferencia de lo observado en el intrón del gen *Dca*, la relación entre sustituciones/*indels* del gen *RC* es muy similar en el *cluster melanogaster* (9 con *D. yakuba* y 6.5 con *D. erecta*) y el *cluster subobscura* (10). Además, el número de sustituciones corregido por la longitud del alineamiento es considerablemente diferente en los dos *clusters* con valores de 0.290 sustituciones/pb (análisis con *D. yakuba*) y 0.209 sustituciones/pb (análisis con *D. erecta*) dentro del *cluster melanogaster* y 0.357 sustituciones/pb en el *cluster subobscura*.

Los *indels* detectados fueron clasificados como inserciones o deleciones. Dentro del *cluster melanogaster*, en el análisis con *D. yakuba* se identificaron 2 *indels*: una deleción de 5pb y una inserción de 1pb. Igualmente, en el análisis con *D. erecta* se observaron dos *indels*: una deleción de 5 pb y una inserción de 3 pb. El *cluster subobscura* presenta 3 *indels*: una deleción de 1pb y dos inserciones de 1pb y 2 pb. De igual forma que en el gen *Dca*, el promedio de la longitud de las deleciones (3.6) es superior al de las inserciones (1.75). Por lo tanto, las deleciones tienden a ser ligeramente más largas que las inserciones.

Tabla III.11. Número de sustituciones e *indels* en el intrón del gen *RC*.

Especies	Alineamiento	Sub	Sub/pb ^a	Indels	Indels/pb ^a	Del	Ins	Del/Ins
Subgrupo <i>melanogaster</i>								
<i>D. yakuba</i> ^b	62	18	0.290	2	0.032	1	1	1
<i>D. erecta</i> ^b	62	13	0.209	2	0.032	1	1	1
Grupo <i>obscura</i>								
<i>D. pseudoobscura</i> ^c	84	30	0.357	3	0.035	1	2	0.5

Sub, sustituciones; Del, deleciones; Ins, inserciones.

^aNúmero total de pares de bases (incluyendo los *gaps*) en el alineamiento.

^bEspecies empleadas como *outgroup* para el subgrupo *melanogaster*.

^cEspecie empleada como *outgroup* para el grupo *obscura*.

Análisis de *Maximum Likelihood (ML)* de la región codificadora del gen *RC*

Al igual que en el gen *Dca*, la divergencia en los genes *RC* ortólogos se analizó con el programa PAML (Yang 1997) utilizando diversos modelos evolutivos y de acuerdo a la filogenia de la figura III.7. Un total de 13 especies fueron empleadas en los diferentes análisis: diez se encuentran incluidas en el subgénero *Sophophora* y tres pertenecen al subgénero *Drosophila*. El gene homólogo *anterior fat body protein* (AFP) de las especies *Calliphora vicina* (*Blue Blowfly*) y *Sarcophoga peregrina* (*Fleshfly*) fue empleado como *outgroup*.

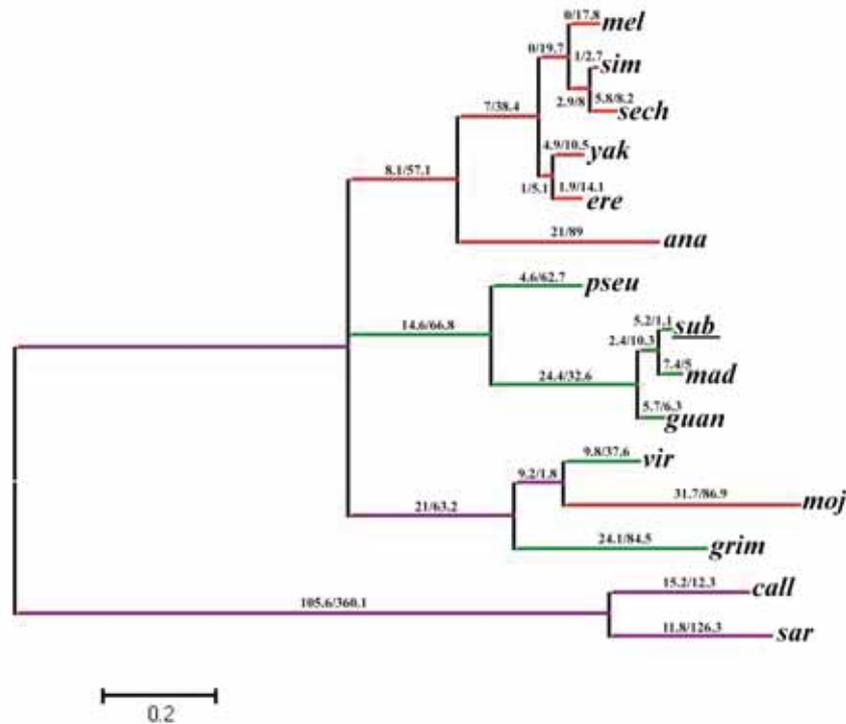


Figura III.7. Filogenia de las 13 secuencias del gen *RC* analizadas de *Drosophila* y las dos especies empleadas como *out-group*. *call*, *Calliphora vicina*; *sar*, *Sarcophoga peregrina*. Los números sobre cada rama corresponden a las estimas por *ML* del número de sustituciones no sinónimas y sinónimas de cada rama de acuerdo al modelo FR. La longitud de las ramas se muestra en proporción a las estimas de divergencia. Nombre de las especies de como en la tabla III.7. El linaje utilizado como *Foreground* en los análisis de *Branch-sites* se presenta subrayado. En rojo, ramas tropicales; en verde, ramas templadas ; en lila, resto de ramas.

En la tabla III. 12 se muestran los valores de *log likelihood* y las estimas de los parámetros bajo los diferentes modelos empleados. Para determinar si los datos observados son compatibles con una presión selectiva homogénea en todas las ramas de la filogenia, el modelo M0 (*one ratio*) que asume las misma ω en todas las ramas del árbol, fue comparado con el modelo *free-ratio* (FR) que permite variación en el valor de

ω en las diferentes ramas de la filogenia. El modelo FR se ajusta significativamente mejor que el modelo M0 ($2 \Delta \ell = 67.083756$, $P = 1.033 \times 10^{-5}$). Sin embargo, las estimas de ω obtenidas con el modelo FR indican que el gen *RC* tiene una restricción funcional elevada a excepción de las ramas que llevan al linaje de *D. subobscura* ($\omega = 1.236$) y la que lleva a *D. virilis* y *D. mojavensis* ($\omega = 1.286$). En la figura III.7 se muestra el número de sustituciones sinónimas y no sinónimas para cada rama de la filogenia calculadas a partir de los parámetros estimados bajo el modelo FR.

Tabla III.12. Valores de *log likelihood* y estimas de los parámetros bajo diferentes modelos

Modelo	f	ℓ	Estimas de los parámetros	Selección positiva
<i>Branch models</i>				
M0	28	- 5387.804394	$\omega = 0.072$, $\kappa = 1.844$	No
FR	53	- 5354.262516	<i>D. subobscura</i> $\omega = 1.236$ <i>D. virilis-D. mojavensis</i> $\omega = 1.286$ $\kappa = 1.815$	Si
<i>BrDsub</i>	29	- 5382.929064	$\omega_0 = 1.514$, $\omega_1 = 0.702$ $\kappa = 1.845$	Si
<i>BrTrop-Temp</i>	30	- 5385.334950	$\omega_{Trop} = 0.064$, $\omega_{Temp} = 0.099$ $\omega_{Res} = 0.061$ $\kappa = 1.84$	No
<i>Site models</i>				
M1	29	- 5337.283429	$\kappa = 2.091$ $p_0 = 0.917$, $\omega_0 = 0.054$ $p_1 = 0.082$, $\omega_1 = 1$	No permitido
M2	31	- 5337.283429	$\kappa = 2.091$ $p_0 = 0.917$, $\omega_0 = 0.053$ $p_1 = 0.082$, $\omega_1 = 1$ $p_2 = 0.000$, $\omega_2 = 114.001$	No
M3 (K=2)	30	- 5264.926122	$\kappa = 1.922$ $p_0 = 0.688$, $\omega_0 = 0.011$ $p_1 = 0.311$, $\omega_1 = 0.237$	No
M3 (K=3)	32	- 5262.790666	$\kappa = 1.926$ $p_0 = 0.628$, $\omega_0 = 0.007$ $p_1 = 0.270$, $\omega_1 = 0.148$ $p_2 = 0.100$, $\omega_2 = 0.376$	No
M7	29	- 5263.819613	$\kappa = 1.928$ $p = 0.276$, $q = 2.954$	No permitido
M8	31	- 5263.819911	$\kappa = 1.928$ $p = 0.276$, $q = 2.954$ $p_0 = 1$ ($p_1 = 0.000$) $\omega_1 = 3.780$	No
<i>Branch-site model A</i>				
<i>Rsub</i>	31	- 5333.818526	$\kappa = 2.093$ $p_0 = 0.678$, $p_1 = 0.058$ $(p_2 + p_3) = 0.263$ $\omega_2 = 3.484$	71E ($P > 0.95$)

f, grados de libertad; ℓ , valor de *log likelihood*; ω , relación d_N/d_S ; κ , relación entre el número de transiciones y transversiones; Trop, ramas tropicales; Temp, ramas templadas; Res, resto de ramas.

Para determinar la posible diferencia en la evolución del gen *RC* en linajes específicos causada por la asociación con la distribución geográfica, nuevamente empleamos un modelo con tres valores de ω diferentes; uno (ω_{Trop}) para los linajes tropicales, uno (ω_{Temp}) para los linajes templados y otro (ω_{Res}) para el resto de ramas. A diferencia de lo observado en el gen *Dca*, este modelo no es significativamente mejor que el modelo M0 ($2 \Delta\ell = 4,938$, $P = 0.084$) indicando que no existe una diferencia significativa entre el valor detectado de $\omega_{\text{Trop}} = 0.065$ y el de $\omega_{\text{Temp}} = 0.099$. Por otro lado, puesto que el test de McDonald y Kreitman (1991) indica una desviación significativa del modelo neutro debido a un exceso de sustituciones no sinónimas fijadas entre *D. subobscura* y *D. guanche* o *D. madeirensis*, se realizaron análisis para determinar si las diferentes estimas de ω a lo largo de las ramas pueden ser causadas por un exceso de sustituciones no sinónimas en el linaje de *D. subobscura*. En este caso fue empleado un modelo con dos ω diferentes; un valor (ω_0) para la rama de *D. subobscura* y otro (ω_1) para el resto de ramas. El anterior modelo (*RCsub*) es significativamente mejor que el modelo M0 ($2 \Delta\ell = 9.750$, $P = 0.0018$).

Los análisis anteriores fueron efectuados asumiendo homogeneidad en la relación ω a lo largo de la secuencia. Sin embargo, este criterio es muy conservativo para detectar selección positiva puesto que la mayoría de aminoácidos de la proteína permanecen muy conservados mientras que algunos pocos pueden ser afectados por la acción de la selección positiva. De esta forma, generalmente el promedio de ω a lo largo de la secuencia no es > 1 , incluso si se ha producido algún evento de adaptación molecular. Por esta razón, se emplearon modelos (*sites models*) (Nilsen y Yang 1998; Yang et al 2000) que permiten variación en el valor de ω entre diferentes posiciones.

En la tabla III.12 se muestran las estimas de máxima verosimilitud (*ML*) y los valores de los parámetros bajo los modelos M1 (*nearly neutral*), M2 (selección), M3 (discreto), M7 (*beta*) y M8 (*beta & ω*). Los resultados obtenidos con los modelos M1, M2 y M3 revelan que el 62-91% de las posiciones analizadas se encuentran bajo fuerte selección purificadora con valores de ω que oscilan entre 0.007 y 0.05. Sin embargo, los valores de ω para el resto de posiciones son más elevados (hasta 0.237), sugiriendo una relajación de la selección (a excepción de los modelos M1 y M2 donde se fija una clase de ω igual a 1). Los modelos M7 y M8 indican que los valores de ω presentan una

distribución beta en forma de L, con la mayoría de aminoácidos conservados o invariables ($p = 0.27$ y $q = 2.95$).

Para establecer la posible existencia de posiciones bajo selección positiva ($\omega > 1$) se compararon los modelos M1 y M2. Los valores idénticos de *ML* obtenidos con los dos modelos indican que no existe ningún codón en particular bajo selección positiva. Por el contrario, los modelos discretos (M3 con $K = 2$ o $K = 3$) se ajustan mejor que el modelo M0. Sin embargo, las estimas de los parámetros bajo el modelo M3 ($K = 3$) no detectan ninguna posición bajo selección positiva ($\omega_2 = 0.376$). Además, este modelo no es significativamente mejor que el modelo más simple M3 ($K = 2$) ($2 \Delta\ell = 4.270$, $P = 0.118$), lo que corrobora la no existencia de posiciones evolucionando con valores de $\omega > 1$. De igual forma, el modelo M8 tampoco identificó ninguna posición bajo selección positiva y no se ajusta significativamente mejor que el modelo alternativo M7.

El modelo alternativo A de *branch-sites* fue implementado para detectar una posible selección positiva afectando solamente a unos pocos aminoácidos en linajes específicos. Este modelo, el cual es una extensión del M1, asume que hay dos clases de sitios ($\omega_0 < 0$ y $\omega_1 = 1$) en todos los linajes y permite que algunos sitios evolucionen bajo selección positiva, incorporando una $\omega_2 > 1$, en el linaje de interés (*foreground branches*). Para el análisis se escogió la rama externa que lleva a *D. subobscura* como *foreground (RCsub)*. Este modelo alternativo A es significativamente mejor que el modelo M1 ($2 \Delta\ell = 6.929$, $P = 0.0312$) y predice una posición (71E) (figura III.8) bajo selección positiva. No obstante, el test 2 (Zhang *et al.* 2005) empleado para evitar la identificación de falsos positivos con el modelo de *branch-sites* no es significativo, cuestionando que la posición 71E haya estado sujeta a selección positiva.

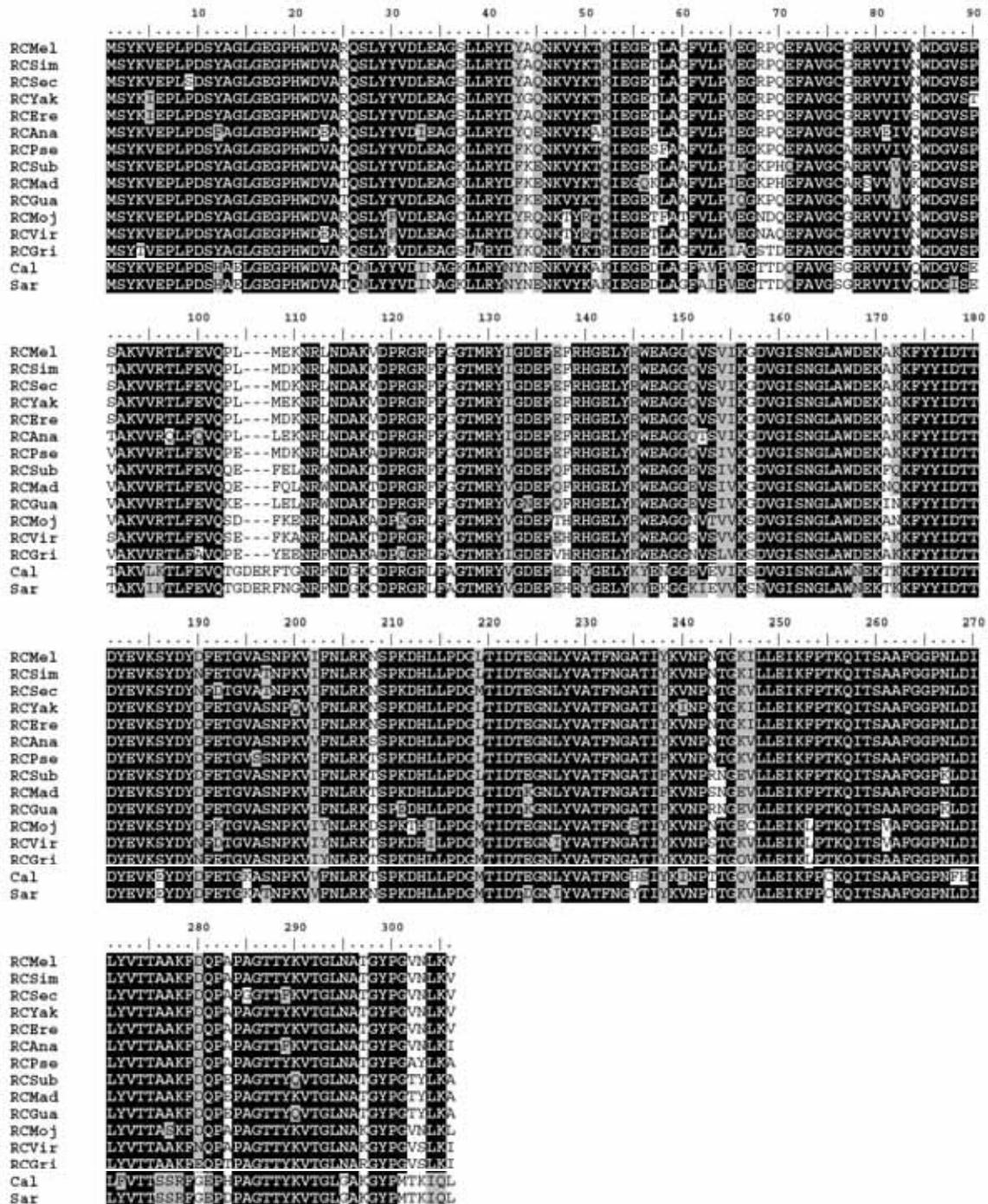


Figura III.8. Alineamiento múltiple de la proteína codificada por el gen RC en las especies analizadas. En negro, aminoácidos idénticos; en gris, aminoácidos con propiedades bioquímicas similares. Mel, *D. melanogaster*; Sim, *D. simulans*; Sec, *D. sechelia*; Yak, *D. yakuba*; Ere, *D. erecta*; Ana, *D. ananassae*; Pse, *D. pseudoobscura*; Sub, *D. subobscura*; Mad, *D. madeirensis*; Gua, *D. guanche*; cal, *Calliphora vicina*; sar, *Sarcophaga peregrina*.

III.2.2 DISCUSIÓN

III.2.2.1 DIVERGENCIA NUCLEOTÍDICA DEL GEN *RC* EN *DROSOPHILA*

La divergencia no sinónima entre *D. melanogaster* y *D. subobscura* del gen *RC* ($K_a = 0.104$) es relativamente elevada en comparación con la detectada, por ejemplo, en los genes *y* ($K_a = 0.068$), *Sxl* ($K_a = 0.045$) y *RpII215* ($K_a = 0.02$), ubicados también en el cromosoma X y analizados previamente entre estas dos especies (Munté et al. 1997, Penalva et al. 1996, Llopard y Agudè 1999). Además el promedio de las estimas de K_a entre las especies del grupo *melanogaster* y *D. subobscura*, *D. madeirensis* o *D. guanche* (0.111, 0.108 y 0.111, respectivamente) fueron aproximadamente el doble del promedio observado entre las especies del grupo *melanogaster* y *D. pseudoobscura* ($K_a = 0.065$). Estos resultados, más que indicar que la proteína RC presenta una baja restricción funcional, sugieren una aceleración en la tasa de fijación de mutaciones no sinónimas en los linajes que conducen a las especies del *cluster subobscura*. De hecho, de acuerdo al *relative-rate* test de Tajima (1993) se ha producido un incremento altamente significativo en la tasa de fijación de las sustituciones no sinónimas en los linajes de *D. subobscura*, *D. madeirensis* y *D. guanche* en relación a *D. pseudoobscura* (tabla III.9).

De acuerdo con el anterior resultado, la relación entre las estimas de divergencia no sinónima y sinónima (K_a/K_s) (tabla III.13) entre las especies del grupo *melanogaster* y *D. pseudoobscura* (desde 0.100 hasta 0.147) son inferiores a los detectados entre las especies del grupo *melanogaster* y *D. subobscura* (desde 0.158 hasta 0.217), *D. madeirensis* (desde 0.152 hasta 0.211) o *D. guanche* (desde 0.156 hasta 0.218) indicando niveles de restricción funcional menores en las tres últimas especies en comparación a *D. pseudoobscura*.

Tabla III.13. Relación K_a/K_s del gen *RC* en *Drosophila*

	<i>mel</i>	<i>sim</i>	<i>sec</i>	<i>yak</i>	<i>ere</i>	<i>ana</i>	<i>pse</i>	<i>sub</i>	<i>mad</i>	<i>gua</i>
<i>mel</i>										
<i>sim</i>	0.061									
<i>sec</i>	0.122	0.297								
<i>yak</i>	0.059	0.112	0.151							
<i>ere</i>	0.031	0.060	0.106	0.129						
<i>ana</i>	0.115	0.143	0.149	0.161	0.158					
<i>pse</i>	0.100	0.100	0.114	0.117	0.108	0.147				
<i>sub</i>	0.158	0.191	0.208	0.210	0.193	0.217	0.172			
<i>mad</i>	0.152	0.183	0.200	0.205	0.188	0.211	0.157	0.826		
<i>gua</i>	0.156	0.188	0.212	0.212	0.208	0.218	0.171	0.306	0.320	

Por otro lado, en el grupo *obscura* se detectaron más diferencias no sinónimas fijadas entre *D. subobscura* y *D. madeirensis* (o *D. guanche*) que las esperadas bajo el modelo neutro como lo revela el test de MK (tabla III.6). Este resultado sugiere la acción de la selección positiva como la causa de la fijación de mutaciones selectivamente ventajosas después de la separación de estos linajes. De hecho, la relación K_a/K_s entre *D. subobscura* y *D. madeirensis* (0.826) y entre *D. subobscura* y *D. guanche* (0.306) es muy superior a la estimada entre *D. pseudoobscura* y las especies del cluster *subobscura* (0.172, 0.157 y 0.171 al considerar *D. subobscura*, *D. madeirensis* o *D. guanche*, respectivamente) indicando una mayor divergencia no sinónima entre las especies paleárticas del grupo *obscura* especialmente en el linaje de *D. subobscura* donde el análisis del PAML indica un valor de $\omega > 1$ probablemente debido a la acción de la selección positiva.

El número de reemplazamientos aminoacídicos fijados entre *D. pseudoobscura* y *D. subobscura*, *D. madeirensis* o *D. guanche* fueron 27, 26 y 28, respectivamente (figura III.9), y en todas las comparaciones aproximadamente el 50% de las sustituciones detectadas son radicales con respecto a la carga. En las comparaciones entre *D. subobscura* y *D. madeirensis* o *D. subobscura* y *D. guanche*, los 12 cambios no sinónimos fijados entre especies producen un total de 11 reemplazamientos aminoacídicos. En la primera comparación, 10 de los cambios observados (90%) son radicales con respecto a la carga (figura III.9). De igual forma, en la segunda comparación 8 de los cambios detectados (72.7%) también producen cambio radical con respecto a la carga.

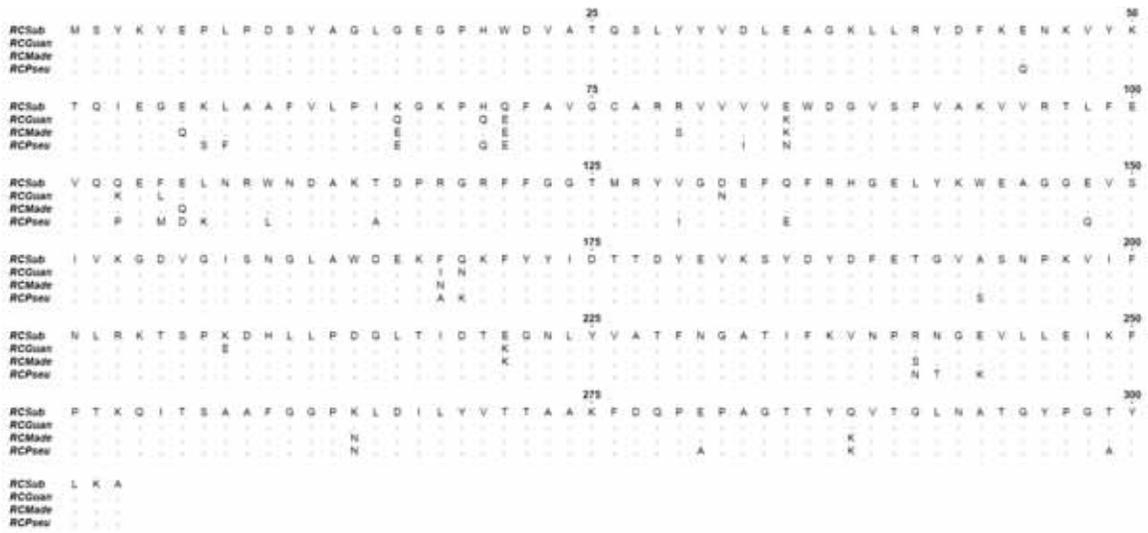


Figura III.9. Alineamiento de la proteína codificada por el gen *RC* en *D. subobscura*, *D. guanche*, *D. madeirensis* y *D. pseudoobscura*. Los puntos (.) indican aminoácidos idénticos a los de la secuencia de *D. subobscura*.

III. 3 LOS GENES PARÁLOGOS *Dca* y *RC*

III.3.1 RESULTADOS

III.3.1.1 EVOLUCIÓN MOLECULAR DE LOS GENES PARÁLOGOS *Dca* Y *RC* EN *D. MELANOGASTER*

En *D. melanogaster* el gen *Dca* se encuentra en la posición 88D2 del cromosoma 3, mientras que el gen *RC* se localiza en el cromosoma X (en la banda 11A). No obstante, la estructura de ambos genes está compuesta por dos exones separados por un pequeño intrón localizado en la misma posición del gen rompiendo el codón en la posición 720 entre la segunda y la tercera base. El nivel de similitud entre las proteínas codificadas por ambos genes es de alrededor del 72.6%. Todas estas características sugieren un origen del gen *Dca* a partir del gen *RC* (o viceversa) por medio de un evento de duplicación génica. En la figura III.9 se muestra el alineamiento de los genes *Dca* y *RC* en *D. melanogaster* y *D. simulans*.

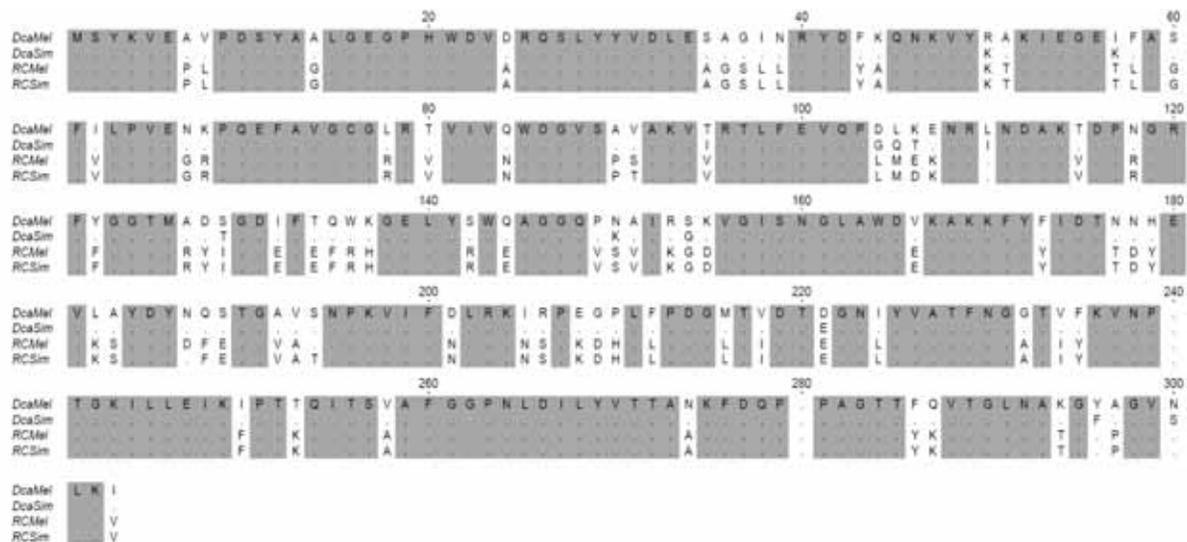


Figura III.10. Alineamiento de la secuencia aminoacídica de las proteínas codificadas por los genes *Dca* y *RC* en *D. melanogaster* y *D. simulans*. Los puntos (.) indican aminoácidos idénticos a la secuencia del gen *Dca* en *D. melanogaster*. En gris se muestran los aminoácidos conservados en ambos genes y especies.

Las estimas de la diversidad nucleotídica sinónima en el gen *Dca* de *D. melanogaster* ($\pi_s = 0.026$) es superior a la detectada en el gen *RC* ($\pi_s = 0.009$). Además, las posiciones no sinónimas de la región codificadora parecen estar bajo diferentes niveles de restricción funcional en los dos genes parálogos. El gen *Dca* presentan un total de 6 polimorfismos aminoacídicos (figura II.2) que dan lugar a 9 variantes proteicas, mientras que en el gen *RC* no se detectaron polimorfismos no sinónimos.

La estima de la divergencia sinónima K_s entre *D. melanogaster* y *D. simulans* presenta mayores niveles en el gen *Dca* ($K_s = 0.139$) en comparación al gen *RC* ($K_s = 0.098$). Los valores detectados en el primer gen pueden considerarse relativamente elevados si se comparan a la estima promedio de la divergencia nucleotídica silenciosa detectada en 19 genes autosómicos ($K_{sil} = 0.108$) entre ambas especies (Begun y Whitley 2000). En el caso del gen *RC*, el valor de K_s observado es inferior al de $K_{sil} = 0.111$ observado en 21 genes ubicados en el cromosoma X, también entre ambas especies (Begun y Whitley 2000).

La evolución de los genes duplicados *Dca* y *RC* fue estudiada por medio de la comparación entre las estimas de divergencia entre los genes parálogos y entre los ortólogos (tabla III.14). Las estimas de la divergencia sinónima (K_s) son superiores que las estimas de divergencia no sinónima (K_a) en todas las comparaciones. Por otro lado, las estimas de divergencia no sinónima entre genes ortólogos difieren considerablemente para ambos genes (0.02 para el gen *Dca* y 0.006 en el gen *RC*), lo que indica que los genes *Dca* y *RC* se encuentran bajo diferentes niveles de restricción funcional. Finalmente, la relación K_a/K_s es también superior entre los genes parálogos que entre los ortólogos. Este resultado podría indicar una selección adaptativa tras la duplicación asociada a la adquisición de una nueva función. No obstante, como la relación K_a/K_s es menor que 1 en todas las comparaciones no puede descartarse una relajación de la restricción funcional como la causa de la fijación de los reemplazamientos aminoacídicos detectados entre los genes *Dca* y *RC*.

El árbol filogenético de los genes duplicados *Dca* y *RC* en *D. melanogaster* y sus respectivas copias en *D. simulans* (figura III.11) revela que ambos genes han evolucionado independientemente desde el evento de duplicación. Entre *D. melanogaster* y *D. simulans* se detectaron un total de 73 reemplazamientos aminoacídicos fijados entre las proteínas DCA y RC (figura III.10).

Tabla III.14. Divergencia de los genes *Dca* y *RC*

Comparaciones	K_s	K_a	K_a/K_s
Entre genes parálogos			
<i>DcaMel-RCSim</i>	0.794	0.232	0.292
<i>RCMel-DcaSim</i>	0.740	0.231	0.312
<i>DcaMel-RCMel</i>	0.770	0.233	0.302
<i>DcaSim-RCSim</i>	0.743	0.230	0.309
Entre genes ortólogos			
<i>DcaMel-DcaSim</i>	0.139	0.020	0.143
<i>RCMel-RCSim</i>	0.098	0.006	0.061

K_s , divergencia sinónima; K_a , divergencia no sinónima;
 K_a/K_s , relación entre la divergencia no sinónima y sinónima;
Mel, *D. melanogaster*; *Sim*, *D. simulans*

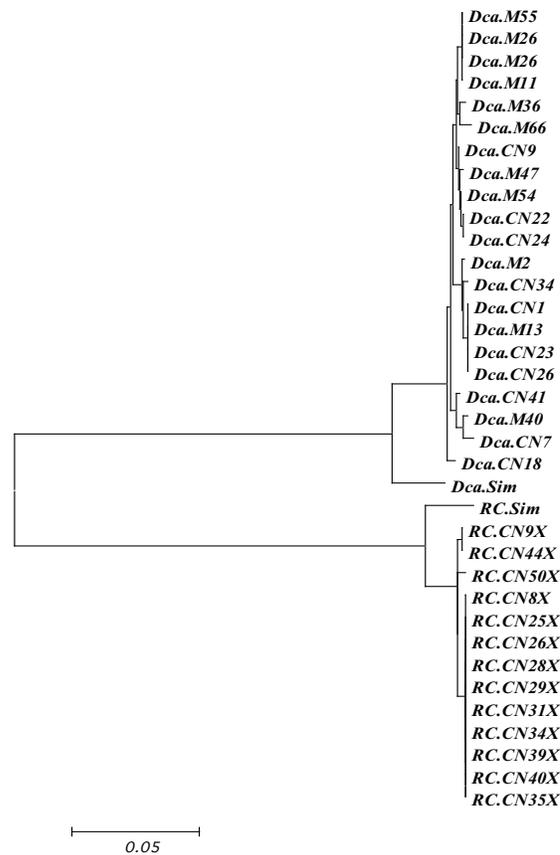


Figura III.11. Árbol filogenético de los genes *Dca* y *RC* en las diferentes líneas analizadas de *D. melanogaster* y en *D. simulans*.

III.3.1.2 EVOLUCIÓN MOLECULAR DE LOS GENES PARALOGOS *Dca* Y *RC* EN *D. SUBOBSCURA*

En *D. subobscura* los genes *Dca* y *RC* se ubican en los cromosomas O y A, respectivamente. De igual forma que lo observado en *D. melanogaster*, la estructura de ambos genes está formada por dos exones separados por un pequeño intrón que se localiza en la misma posición rompiendo el codón en la posición 720 entre la segunda y la tercera base. En esta especie, la similitud entre las proteínas codificadas por ambos genes es de alrededor del 71.6%, muy similar a la hallada en *D. melanogaster*.

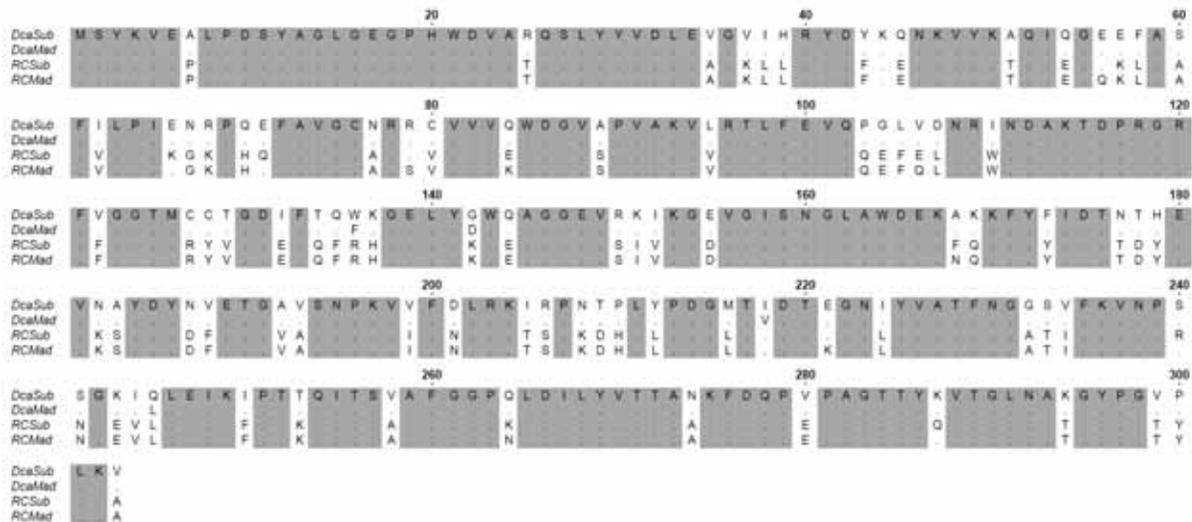


Figura III.12. Alineamiento de la secuencia aminoacídica de las proteínas codificadas por los genes *Dca* y *RC* en *D. subobscura* y *D. madeirensis*. Los puntos (.) indican aminoácidos idénticos a la secuencia del gen *Dca* de *D. subobscura*. En gris se muestran los aminoácidos conservados en ambos genes y especies.

Los niveles de la diversidad nucleotídica en la región codificadora de los genes *Dca* y *RC* en *D. subobscura* indican diferentes niveles de restricción funcional. De hecho, el gen *RC* presenta unas estimas de la diversidad nucleotídica inferiores a las del gen *Dca* para las posiciones sinónimas y especialmente para las no sinónimas ya que no se detectó ningún polimorfismo de reemplazamiento (tablas II.4 y III.4). Esta diferencia en los niveles de polimorfismo no sinónimo se ve reflejada en el mayor número de variantes proteicas que presenta el gen *Dca* (10) en comparación a la única detectada en el gen *RC*.

La divergencia sinónima de ambos genes (K_s) se estimó en la comparación entre *D. subobscura* y *D. madeirensis*. La estima en el gen *Dca* ($K_s = 0.050$) es el doble de la detectada en el gen *RC* ($K_s = 0.023$). Además, la divergencia sinónima del gen *Dca* es relativamente elevada si se compara con las estimas previamente publicadas de los genes autonómicos *rp49* ($K_{sil} = 0.011$; Rozas y Aguadé 1994) y *Acp70A* ($K_{sil} = 0.027$; Cirera y Aguadé 1998) entre *D. subobscura* y *D. madeirensis*. No obstante, el gen *RC* presenta unos niveles de divergencia sinónima muy similares a los de divergencia silenciosa detectados entre estas dos especies en el gen *y* ($K_{sil} = 0.026$) ubicado también en el cromosoma X.

Las estimas de divergencia sinónima (K_s) son superiores que las estimas de divergencia no sinónima (K_a) tanto entre los genes parálogos como entre los genes ortólogos (tabla III.15). De la misma forma que lo observado entre *D. melanogaster* y *D. simulans*, las estimas de la divergencia no sinónima entre *D. subobscura* y *D. madeirensis* difiere entre copias (0.008 para el gen *Dca* y 0.019 en el gen *RC*), lo que indica que los genes *Dca* y *RC* en estas especies también se encuentran bajo diferentes niveles de restricción funcional. De hecho, en el gen *Dca* la relación entre la divergencia no sinónima y sinónima (K_a/K_s) entre *D. subobscura* y *D. madeirensis* es bastante inferior que 1, mientras que el gen *RC* presenta un valor de $K_a/K_s = 0.826$ (tabla III.15). La relación K_a/K_s varía desde 0.736 hasta 0.951 con un valor promedio de 0.832 al emplear cualquiera de las 12 líneas de *D. subobscura* secuenciadas del gen *RC*. Por lo tanto, esto podría indicar la acción de la selección positiva en el gen *RC* después de la separación de estas dos especies. No obstante, la relajación de la selección natural como la causa de la fijación de los reemplazamientos aminoacídicos no puede ser descartada ya que la relación K_a/K_s permanece inferior a 1.

Al igual que en el análisis con *D. melanogaster* y *D. simulans*, el árbol filogenético de los genes *Dca* y *RC* en *D. subobscura* y *D. madeirensis* revela que ambos genes han evolucionado independientemente desde su duplicación (figura III.13). No obstante, en el caso del gen *Dca*, *D. madeirensis* se agrupa con las líneas analizadas de *D. subobscura* en lugar de ubicarse como especie *outgroup*. La proteína codificada por los genes *Dca* y *RC* en las anteriores especies presentan 75 reemplazamientos aminoacídicos fijados (figura III.12)

Tabla III.15. Divergencia de los genes *Dca* y *RC*

Comparaciones	K_s	K_a	K_a/K_s
Entre genes parálogos			
<i>DcaSub-RCMad</i>	0.516	0.222	0.430
<i>RCSub-DcaMad</i>	0.543	0.221	0.419
<i>DcaSub-RCSub</i>	0.508	0.220	0.433
<i>DcaMad-RCMad</i>	0.552	0.223	0.404
Entre genes ortólogos			
<i>DcaSub-DcaMad</i>	0.050	0.008	0.160
<i>RCSub-RCMad</i>	0.023	0.019	0.826

K_s , divergencia sinónima; K_a , divergencia no sinónima;
 K_a/K_s , relación entre la divergencia no sinónima y sinónima.
Sub, *D. subobscura*; *Mad*, *D. madeirensis*.

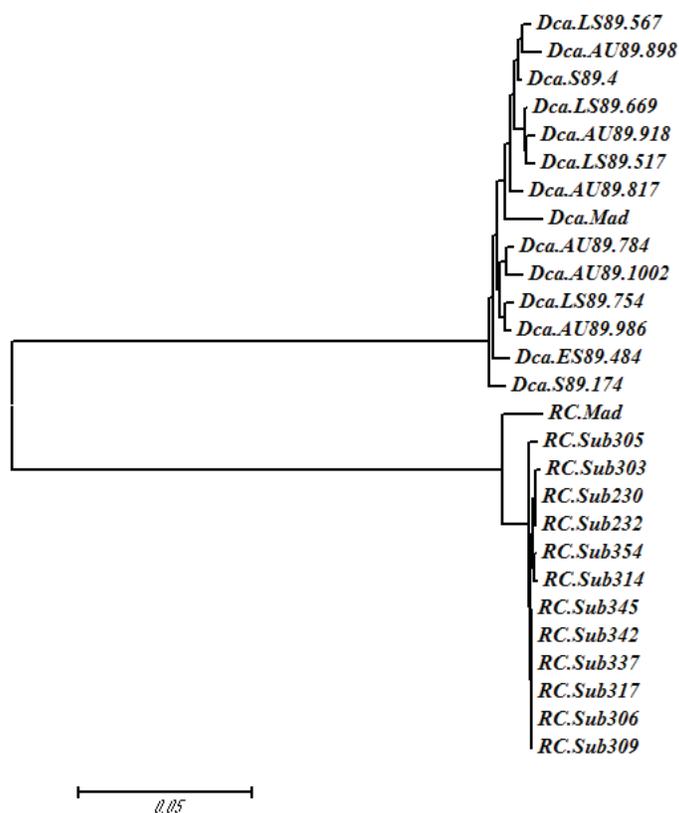


Figura III.13. Árbol filogenético de los genes *Dca* y *RC* en las diferentes líneas analizadas de *D. subobscura* y *D. madeirensis*.

III.3.1.3 ANÁLISIS DE *MAXIMUM LIKELIHOOD* (*ML*) EN LA REGIÓN CODIFICADORA DE LOS GENES PARÁLOGOS *DCA* Y *RC* (*DR*)

Los resultados de los modelos FR y M0 para el conjunto de los genes parálogos *Dca* y *RC* (*DR*) son semejantes a los obtenidos previamente para cada copia individualmente. El modelo FR es significativamente mejor que el modelo nulo M0 ($2 \Delta\ell = 160.385$, $P = 1.742 \times 10^{-12}$) e indica que en el gen *RC* la rama que lleva al linaje de *D. subobscura* presenta un valor de $\omega > 1$.

Para identificar la acción de la selección positiva en ramas específicas del conjunto *DR* consideramos un modelo con dos valores de ω diferentes, un valor (ω_1) para ramas concretas y otro valor (ω_0) para el resto de ramas de la filogenia. Los valores de ω_1 se fijaron alternativamente en la rama interna previa a la divergencia del gen *Dca* (*dDca*) o del gen *RC* (*dRC*) o en la rama externa que conduce al gen *RC* de *D. subobscura* (*RCSUB*) (figura III.14). Este modelo se ajusta significativamente mejor que el modelo M0 en los tres casos (para *dDca*, $2 \Delta\ell = 16.692$, $P = 4.394 \times 10^{-5}$; para *dRC*, $2 \Delta\ell = 3.655$, $P = 0.056$; para *RCSUB*, $2 \Delta\ell = 7.684$, $P = 0.005$).

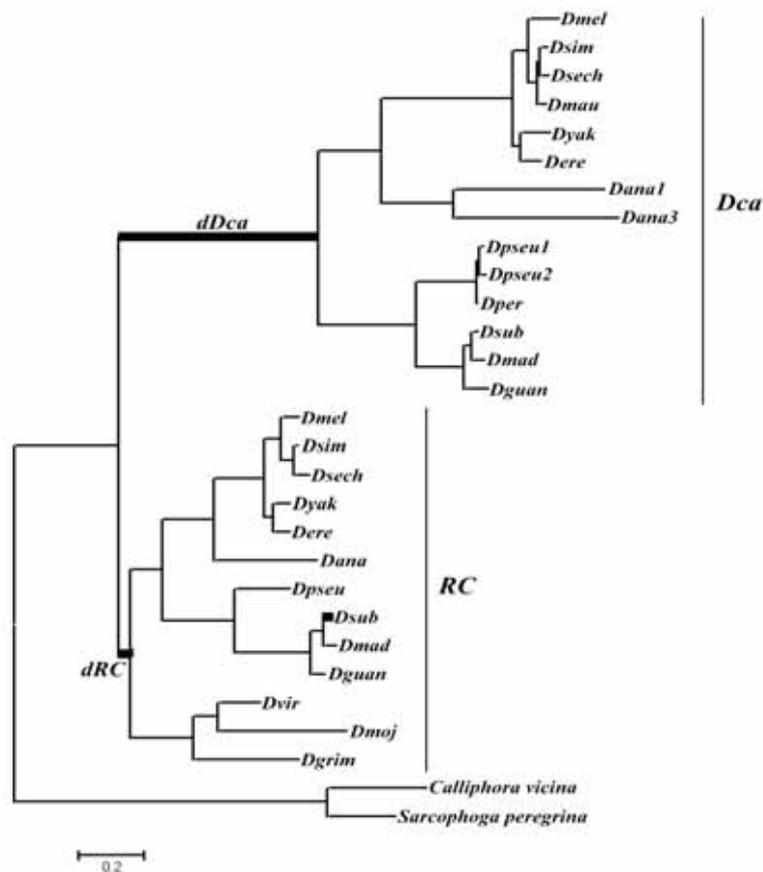


Figura III.14 Ramas usadas como *foreground* en los análisis de *branch-sites*

Los resultados de las estimas de *ML* al emplear los modelos M1 (*nearly neutral*), M2 (selección), M3 (discreto), M7 (*beta*) y M8 (*beta & ω*) se muestran en la tabla III.16.

Tabla III.16. Valores de *log likelihood* y estimas de los parametros bajo difrentes modelos en los genes parálogos *Dca* y *RC* (*DR*)

Modelo	f	ℓ	Estimas de los parámetros	Selección positiva
<i>Branch models</i>				
M0	57	-8931.652619	$\omega = 0.088, \kappa = 1.623$	No
FR	111	-8851.459780	<i>D.subobscura</i> $\omega = 1.132$ $\kappa = 1.519$	Si
<i>BrDsub</i>	58	-8927.810382	$\kappa = 1.623$ $\omega_2 = 0.087, \omega_1 = 1.359$	Si
<i>BrDca</i>	58	-8923.306179	$\kappa = 1.605$ $\omega_2 = 0.083, \omega_1 = 0.368$	No
<i>BrDRC</i>	58	-8929.824952	$\kappa = 1.624$ $\omega_2 = 0.081, \omega_1 = 999.000$	No
<i>Sites models</i>				
M1	58	-8849.326418	$\kappa = 1.781$ $p_0 = 0.919, \omega_0 = 0.076$ $p_1 = 0.081, \omega_1 = 1$	No permitido
M2	60	-8849.326468	$\kappa = 1.781$ $p_0 = 0.919, \omega_0 = 0.076$ $p_1 = 0.081, \omega_1 = 1$ $p_2 = 0.000, \omega_2 = 42.305$	No
M3 (K=2)	59	-8724.388702	$\kappa = 1.671$ $p_0 = 0.588, \omega_0 = 0.016$ $p_1 = 0.411, \omega_1 = 0.218$	No
M3 (K=3)	61	-8705.378453	$\kappa = 1.683$ $p_0 = 0.522, \omega_0 = 0.010$ $p_1 = 0.404, \omega_1 = 0.155$ $p_2 = 0.073, \omega_2 = 0.484$	No
M7	58	-8708.726227	$\kappa = 1.672$ $p = 0.420, q = 3.441$	No permitido
M8	60	-8705.047905	$\kappa = 1.672$ $p = 0.463, q = 4.314$ $p_0 = 0.988 (p_{11} = 0.011)$ $\omega_1 = 1.046$	No
<i>Branch-sites model A</i>				
Rama <i>RCsub</i>	60	-8845.924182	$\kappa = 1.782$ $p_0 = 0.566, p_1 = 0.050$ $(p_2 + p_3) = 0.382$ $\omega_2 = 2.695$	66E, 71E, 172A ($P > 0.95$)
Rama <i>dDca</i>	60	-8832.580689	$\kappa = 1.765$ $p_0 = 0.785, p_1 = 0.065$ $(p_2 + p_3) = 0.148$ $\omega_2 = 3.676$	60S, 136L, 138T, 141K, 186L, 278N ($P > 0.95$)
Rama <i>dRC</i>	60	-8848.469667	$\kappa = 1.765$ $p_0 = 0.000, p_1 = 0.000$ $(p_2 + p_3) = 1$ $\omega_2 = 41.897$	No

f, grados de libertad; ℓ, valor de *log likelihood*; ω, relación d_N/d_S ; κ, relación entre el número de transiciones y transversiones.

Los análisis realizados con los modelos M1, M2 y M3 ($K=2$ y $K=3$) indican que el 52-91% de las posiciones analizadas se encuentran bajo una fuerte selección purificadora con valores de ω que van desde 0.01 hasta 0.07. Sin embargo, los valores de ω para el resto de posiciones son más elevados (hasta 0.484), sugiriendo una relajación de la selección (a excepción de los modelos M1 y M2 donde se fija una clase de $\omega = 1$). Las estimas de los parámetros obtenidos con los modelos M7 y M8 sugieren que los valores de ω presentan una distribución *beta* en forma de *L*, con la mayoría de aminoácidos muy conservados o poco variables ($p \sim 0.4$ y $q \sim 3.7$).

La diferencia entre los dos modelos M1 y M2 no es significativa debido a que las estimas de *ML* son muy similares y por lo tanto no existen evidencias de selección positiva en ningún codón en particular. Por el contrario, los modelos discretos (M3 con $K = 2$ o $K = 3$) se ajustan mejor que el modelo M0. Adicionalmente, el modelo M3 ($K = 3$) es significativamente mejor que el modelo más simple M3 ($K = 2$) ($2 \Delta\ell = 38.020$, $P = 5.545 \times 10^{-9}$). Sin embargo, las estimas de los parámetros bajo este modelo no detectan ninguna posición bajo selección positiva ($\omega_2 = 0.155$).

El modelo alternativo A de *branch-sites* implementado con el propósito de determinar la existencia de sitios bajo selección positiva en linajes específicos es significativamente mejor que el modelo M1 únicamente en las ramas *dDca* ($2 \Delta\ell = 33.491$, $P = 5.338 \times 10^{-8}$) y *RCSUB* ($2 \Delta\ell = 6.804$, $P = 0.033$). Este modelo identificó 6 (60S, 136I, 138T, 141K, 186L y 278N) y 3 (66E, 71E y 172A) (figura III.15) posiciones con una alta probabilidad posterior de evolucionar bajo selección positiva en las ramas *dDca* y *RCSUB*, respectivamente. El test 2 (Zhang et al. 2005) se empleó para evitar la identificación de falsos positivos con el modelo de *branch-sites*. Este test es significativo en la rama *dDca* ($2 \Delta\ell = 4.391$, $P = 0.036$) pero no en *RCSUB* ($2 \Delta\ell = 0.190$, $P = 0.662$). Por lo tanto, a pesar que los modelos de *branch-sites* predicen posiciones seleccionadas positivamente en linajes específicos, no se puede descartar que el valor de $\omega_2 > 1$ en el linaje *RCSUB* sea causado por la relajación de la restricción funcional ($\omega_2 \approx 1$), mientras que la selección positiva puede haber actuado en la rama *dDca*.

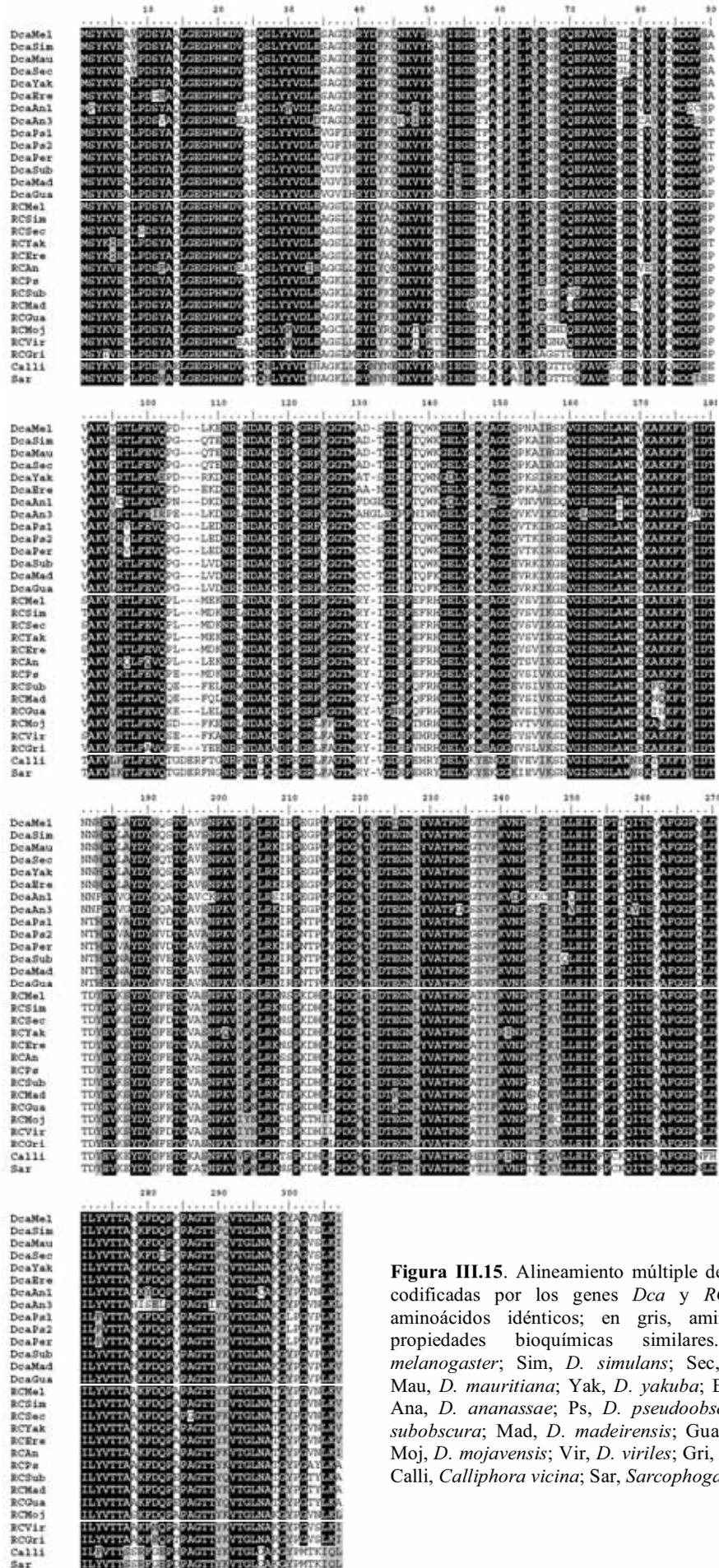


Figura III.15. Alineamiento múltiple de las proteínas codificadas por los genes *Dca* y *RC*. En negro, aminoácidos idénticos; en gris, aminoácidos con propiedades bioquímicas similares. Mel, *D. melanogaster*; Sim, *D. simulans*; Sec, *D. sechelia*; Mau, *D. mauritiana*; Yak, *D. yakuba*; Ere, *D. erecta*; Ana, *D. ananassae*; Ps, *D. pseudoobscura*; Sub, *D. subobscura*; Mad, *D. madeirensis*; Gua, *D. guanche*; Moj, *D. mojavensis*; Vir, *D. viriles*; Gri, *D. grimshawi*; Calli, *Calliphora vicina*; Sar, *Sarcophoga peregrina*.

III.3.2 DISCUSIÓN

III.3.2.1 EVOLUCIÓN MOLECULAR Y ANÁLISIS DE *ML* DE LOS GENES *DCA* Y *RC* EN *DROSOPHILA*

Las regiones ortólogas de los genes *Dca* y *RC* fueron identificadas en todas las especies del subgenero *Sophophora*, incluyendo a *D. willistoni* en la que se detectaron dos copias del gen *Dca* y una del gen *RC*. Sin embargo, esta última especie no fue incluida en los análisis de *ML* debido al cambio que presenta en la composición de bases y en la preferencia de codones (Vicario et al. 2007; Singh et al. 2006; *Drosophila* 12 Genome Consortium 2007). Por el contrario, en las tres especies incluidas en el subgénero *Drosophila* (*D. grimshawi*, *D. virilis* y *D. mojavensis*) se detectó una única región con altos niveles de sintenia con el gen *RC* ubicado en el cromosoma X. Por otro lado, las especies *Calliphora vicina* (*Blue Blowfly*) y *Sarcophoga peregrina* (*Fleshfly*) empleadas como *outgroup*, presentan una única región de similitud a los genes *Dca* y *RC*, siendo ésta más elevada cuando dicha región se compara con el gen *RC*. Estos datos sugieren que el evento de duplicación ocurrió después de la separación de los subgéneros *Drosophila-Sophophora* pero antes de la radicación de las especies del subgénero *Sophophora* (figura III.14).

Además del alto grado de similitud de las secuencias aminoacídicas de las proteínas *DCA* y *RC*, la estructura de los genes que las codifican con un pequeño intron de fase 2 también está conservada en todas las especies de *Drosophila*. El mantenimiento de estos genes parálogos durante un periodo de tiempo tan largo sugiere la acción de un mecanismo de preservación. Existen dos principales teorías para el mantenimiento de las dos copias de un gen duplicado: subfuncionalización y neofuncionalización (Ohno 1970; Clark 1994; Walsh 1995). La subfuncionalización de genes duplicados es un proceso importante en el modelo DDC (*Duplication-Degeneration-Complementation*), en el cual las funciones de un gen ancestral son divididas entre los genes duplicados que funcionalmente se complementan mutuamente. Por otra parte, la neofuncionalización de genes duplicados resulta en la adquisición de una nueva función en uno de los genes, mientras que la función ancestral es preservada por otro gen.

La proteína de la familia SMP-30 de las especies *S. peregrina* y *C. vicina* empleadas como *outgroup* en los análisis fue denominada *anterior fat body protein* (AFP) debido a que su expresión se encuentra altamente restringida al citoplasma de los trofocitos (células grasas) de la región anterior del cuerpo graso de la larva (Nakajima y Natori 2000; Hansen et al. 2002). La AFP de *S. peregrina* esta constituida por 306 aminoácidos y su secuencia presenta un 75% y 61% de residuos idénticos a los de la proteína codificada por los genes *RC* y *Dca* de *D. melanogaster*, respectivamente. El mayor grado de similitud de AFP con la proteína *RC* de las especies de *Drosophila* analizadas, sugiere que el gen AFP es homólogo del gen *RC* en *Drosophila* y que el gen *Dca* surgió tras un evento de duplicación.

La posible función del gen AFP en *C. vicina* es la incorporación (endocitosis) de proteínas de almacenamiento (hexamerinas) desde la hemolinfa al cuerpo graso por medio de la interacción con el receptor de hexamerinas (Hansen et al. 2002) durante el proceso de metamorfosis. La interacción entre estas dos moléculas solamente puede ocurrir en la parte anterior del cuerpo graso, un tejido que se degrada rápidamente poco después de la formación de la pupa. Por otro lado, el gen AFP de *S. peregrina* no parece exhibir una fuerte afinidad con el calcio, en contraste con el gen *RC* de los mamíferos, el cual se asume que ha derivado del mismo ancestro común (Nakajima y Natori 2000).

Por lo tanto, es posible que el gen *RC* en las especies del *Drosophila* mantenga la función ancestral, mientras que el gen *Dca* pudo haber adquirido una nueva función después del evento de duplicación. De hecho, Goto (2000) ha sugerido que la posible función del gen *Dca* en *Drosophila*, al igual que la descrita para el gen SMP30 de los mamíferos, es el mantenimiento de los niveles de calcio (Ca^+) en el citosol. El aumento detectado en los niveles de la expresión del gen *Dca* después de la exposición a 15°C en individuos de *Drosophila* podría indicar la importancia de este gen en el mantenimiento de los niveles de Ca^+ a bajas temperaturas y en la tolerancia al frío en *Drosophila*. No obstante, para poder establecer la posible diferenciación funcional de los genes *Dca* y *RC* es fundamental determinar la función y el patrón de expresión que presentan estos genes en *Drosophila*.

Los análisis de *ML* muestran que las estimas de d_N/d_S (ω) en la región codificadora de los genes *Dca* y *RC* son significativamente heterogéneas y bastante variables entre los diferentes linajes analizados, indicando que la evolución de estos genes es incompatible con el modelo neutral. No obstante, ambos genes presentan

valores de $d_N/d_S < 1$ (a excepción de las ramas *RCSUB* y la que lleva a *D. virilis* y *D. mojavensis* del gen *RC*) sugiriendo que han evolucionado bajo selección purificadora.

El modelo empleado para determinar la posible diferencia en la evolución de los genes *Dca* y *RC* en linajes específicos causada por la asociación con la distribución geográfica, indica que el gen *Dca* presenta una diferencia significativa del patrón ω en los linajes de las especies tropicales respecto a los de las especies templadas con niveles de restricción funcional mayores en estas últimas. Por el contrario, esta diferencia entre linajes tropicales y templados no se detecta en el gen *RC*. Este resultado estaría de acuerdo con la implicación de la tolerancia al frío del gen *Dca* pero no del gen *RC*.

En las ramas que llevan hacia los linajes tropicales del gen *Dca*, el modelo A de *Branch-sites* identificó un total de 16 reemplazamientos. Uno de estos reemplazamientos (104D) pudo haber evolucionado por selección positiva (figura III.15). No obstante, el test 2 (Zhang et al. 2005) empleado para evitar la clasificación falsa de posiciones bajo selección positiva no descarta la relajación de la restricción funcional como la causa del alto valor de ω detectado en el anterior reemplazamiento. Por el contrario, el test 2 sugiere que los dos reemplazamientos identificados (140W y 146S) en las ramas templadas del gen *Dca* con valores elevados de ω han evolucionado por selección positiva.

El análisis conjunto de los genes parálogos y ortólogos *Dca* y *RC* permitió detectar que en las ramas *dDca* y *RCSUB* el modelo A de *branch-sites* es significativamente mejor que el modelo M1. De los 41 reemplazamientos detectados en *dDca*, 6 han evolucionado por selección positiva (60S, 136I, 138T, 141K, 186L y 278N) (figura III.15). De éstos, 3 son radicales con respecto a la carga, 3 con respecto a la polaridad y 1 es radical para ambos criterios. El test 2 (Zhang et al. 2005) permite concluir que la selección positiva pudo haber actuado sobre estos cambios aminoacídicos en la rama *dDca* mientras que no descarta la relajación de la restricción funcional como la causa de los altos valores de ω detectados en la rama *RCSUB*.

A pesar que en este trabajo se han identificado posiciones aminoacídicas que han evolucionado bajo selección positiva y que pueden causar características diferenciales en los genes *Dca* y *RC* después del evento de duplicación, es indispensable información adicional tanto de la estructura como de la función de las proteínas codificadas por estos genes para demostrar el carácter adaptativo de los aminoácidos identificados bajo selección positiva por medio de los análisis de *ML*. Además, como se mencionaba anteriormente, el análisis de las regiones reguladoras y el estudio de posibles diferencias

en los niveles y el patrón de expresión de los genes *Dca* y *RC* son fundamentales para poder comprender mejor el papel de la selección natural en la evolución molecular de estos genes en *Drosophila*.

CAPÍTULO IV
EL GEN CG6296

IV.1 POLIMORFISMO NUCLEOTÍDICO DEL GEN CG6296

IV.1.1 RESULTADOS

IV.1.1.1 POLIMORFISMO NUCLEOTÍDICO DEL GEN CG6296 EN *D. MELANOGASTER*

La variación nucleotídica de la región CG6296 fue analizada en 19 líneas de *D. melanogaster* procedentes de dos poblaciones naturales españolas (Montemayor y Sant Sadurní d'Anoia). El fragmento estudiado de aproximadamente 2.7 kb incluye alrededor de 2.1 kb de la región codificadora, así como parte de las regiones 5' y 3' flanqueantes con una longitud aproximada de 0.4 y 0.2 kb, respectivamente.

Diferenciación genética entre poblaciones

La diferenciación genética entre las poblaciones de Montemayor y Sant Sadurní d'Anoia fue estimada usando el estadístico F_{st} (Hudson et al 1992). El valor observado al comparar las dos poblaciones analizadas ($F_{st} = 0.0069$) no difiere significativamente del esperado bajo el modelo neutro en muestras aleatorias de una población panmítica sin subdivisión (figura IV.1). Por lo tanto, ambas poblaciones se consideraron conjuntamente en los análisis posteriores.

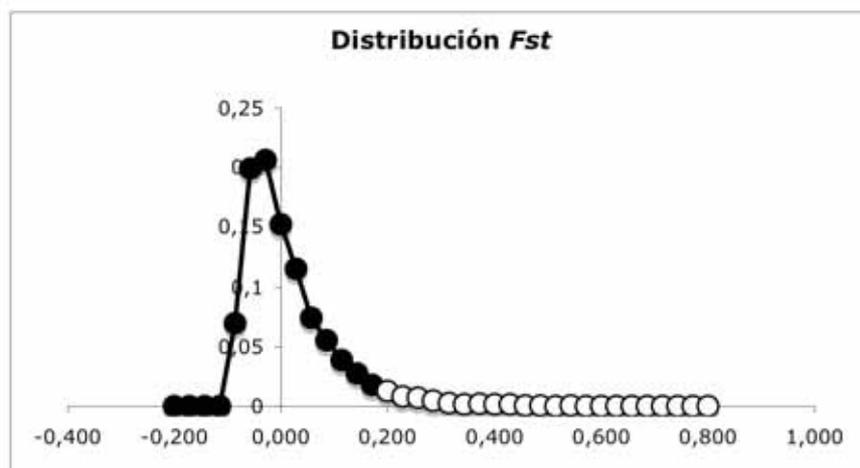


Figura IV.1. Distribución del estadístico F_{st} en una población panmítica de 19 individuos (ver materiales y métodos).

Análisis del polimorfismo nucleotídico

El alineamiento múltiple de las 19 secuencias de *D. melanogaster* presenta un total de 2814 posiciones. En dicho alineamiento se detectaron tanto polimorfismos por cambio nucleotídico como polimorfismos por inserción/delección. Al eliminar las posiciones que presentan *gaps* en una o varias de las líneas analizadas el número de posiciones se reduce a 2770. En total se identificaron 84 posiciones polimórficas por cambio nucleotídico y 5 polimorfismos por inserción/delección de nucleótidos.

En la región 5' flanqueante se detectaron 16 polimorfismos por cambio nucleotídico, de los que 2 presentan variantes únicas. También se detectaron 4 polimorfismos por inserción/delección. La región codificadora presenta un total de 56 polimorfismos por cambio nucleotídico, de los cuales 29 son sinónimos, 27 no sinónimos y 17 presentan variantes únicas. Además de los polimorfismos por cambio nucleotídico, la región codificadora presenta un polimorfismo por inserción/delección de 3 nucleótidos. En la región 3' flanqueante se detectaron 7 polimorfismos por cambio nucleotídico, de los cuales 2 presentan variantes únicas (figura IV.2).

CN1	T	G	G	C	T	A	A	T	G	T	T	d3	174	152	114	113	79	82	55	22	21	15	11
CN5	C	A	A	T	d32	A	d5	A	T	T	T	C	176	184	188	190	277	365	365	468	478	511	558
CN6																							
CN7	C	A			d32	A	d5	A	d3														
M2																							
M11																							
M22																							
M26																							
M54	A	T			T	d32	A	d5	A														
CN23	A	T			T	T	T	C															
CN36	A	T			T	d32	A	d5	A														
M13	A	T			T	d32	A	d5	A														
M40																							
M36																							
M55	A	T																					
CN9	A	T	A		T	d32	A	d5	A														
M66	A	T	A	C	T	d32	A	d5	A														
CN34																							
<i>D. simulans</i>																							

CN1	G	G	T	C	C	A	T	G	G	A	T	G	G	A	T	G	T	T	G	T	C	G	G	T	A	G	C	A	C	2769
CN5																														2825
CN6																														2804
CN7																														2802
CN22																														2594
M2																														2584
M11																														2578
M22																														2547
M26																														2480
M54																														2414
CN23																														2313
CN36																														2250
M13																														2246
M40																														2245
M36																														2228
M55																														2221
CN9																														2206
M66																														2192
CN34																														2144
<i>D. simulans</i>																														2127

Figura IV.2. Polimorfismo nucleotídico del gen CG6296 en las 19 líneas analizadas de *D. melanogaster* de las poblaciones de Montemayor (M) y Sant Sadurní d'Anoia (CN). Las posiciones polimórficas están numeradas según el alineamiento múltiple de las secuencias de *D. melanogaster*. Las posiciones polimórficas en la región codificadora se indican sombreadas en gris. NS, polimorfismo no sinónimo. Los puntos (.) indican nucleótidos idénticos a los de la secuencia de referencia y los guiones (-) posiciones con *gaps*. Las deleciones de nucleótidos respecto a la secuencia de referencia así como su longitud se indican con d#. La última fila muestra la información en *D. simulans* para las posiciones identificadas como polimórficas en *D. melanogaster*.

Análisis del polimorfismo aminoacídico

El alineamiento múltiple de las secuencias aminoacídicas de la proteína codificada por el gen CG6296 en las 19 líneas de *D. melanogaster* analizadas se presenta en la figura IV.3. Este alineamiento comprende un total de 676 posiciones. Se observaron 27 polimorfismos de reemplazamiento de los cuales 11 presentan variantes únicas. Es de destacar que 9 de los 11 variantes únicas están presentes en una sola línea (M66).

Los polimorfismos aminoacídicos detectados causan cambios tanto conservativos como radicales dependiendo de si alteran o no las propiedades físico-químicas de los aminoácidos. Con respecto a la carga, se detectaron 20 cambios conservativos (por ejemplo, los cambios ácido glutámico/ácido aspártico (Glu/Asp), valina/isoleucina (Val/Ile) y leucina/serina en las posiciones 600, 1350 y 2206, respectivamente) y 7 radicales (por ejemplo, los cambios ácido glutámico/lisina (Glu/Lys) y ácido glutámico/valina (Glu/Val) en las posiciones 790 y 1999, respectivamente). Con respecto a la polaridad, se observaron un total de 18 cambios conservativos (por ejemplo: isoleucina/metionina (Ile/Met), leucina/valina (Leu/Val) y serina/asparagina (Ser/Asn) en las posiciones 768, 1575 y 2014 respectivamente) y 9 radicales (por ejemplo: treonina/alanina (Thr/Ala) y alanina/serina (Ala/Ser) en las posiciones 1605 y 2127, respectivamente). Al considerar la secuencia de *D. simulans*, 10 de los 11 aminoácidos que se encuentran segregando a baja frecuencia en las posiciones polimórficas son derivados.

Además de los polimorfismos de reemplazamiento, la región codificadora del gen CG6296 presenta un polimorfismo por inserción /delección. Así, la línea M40 presenta una delección de un aminoácido con respecto a la secuencia de referencia. Dicha delección está ubicada hacia el extremo carboxi-terminal de la proteína.

Diversidad haplotípica

Las 19 líneas de *D. melanogaster* analizadas pueden agruparse en un total de 15 haplotipos diferentes si se consideran todos los polimorfismos por cambio nucleotídico detectados en la región estudiada. La diversidad haplotípica detectada es $H_d = 0.959$.

Asimismo, en la región codificadora se detectaron un total de 15 haplotipos que representan 8 variantes proteicas. En la figura IV.3 se observa que la variante P1 presenta una frecuencia del 52.6% (10 líneas) y la variante P3 del 10.5% (2 líneas). El resto de variantes (P2, P4-P9) se encuentran en una sola línea y se diferencian por más de una sustitución aminoacídica.



Figura IV.3. Polimorfismo aminoacídico de la proteína codificada por el gen CG6296 en las 19 líneas de *D. melanogaster* analizadas de las poblaciones de Montemayor (M) y Sant Sadurní d'Anoia (CN). En la parte superior se indican las posiciones de los polimorfismos nucleotídicos que causan los cambios aminoacídicos de acuerdo a la figura 19. Los puntos (.) indican aminoácidos idénticos a los de la secuencia de referencia y los guiones (-) posiciones con *gaps*. P# indica la variante proteica. La última fila muestra la información en *D. simulans* para las posiciones identificadas como polimórficas en *D. melanogaster*.

Desequilibrio de ligamiento y recombinación

El grado de desequilibrio de ligamiento se analizó empleando las posiciones informativas. De las 1830 comparaciones realizadas entre los 61 polimorfismos detectados, 554 son significativas por el test χ^2 (30.3%) y 68 con la corrección de Bonferroni (3.7%). Se detectaron un total de 13 eventos de recombinación (Rm) en la región analizada de acuerdo al test de los cuatro gametos de Hudson y Kaplan (1985).

Diversidad nucleotídica

En la tabla IV.1 se muestran las estimas del polimorfismo nucleotídico obtenidas para diferentes tipos de posiciones. La diversidad nucleotídica para toda la región analizada (π_{total}) es 0.009 y ligeramente inferior a la detectada para las posiciones

silenciosas ($\pi_{\text{sil}} = 0.015$). Dentro de la región codificadora, la estima para las posiciones sinónimas (π_s) es 0.019, mientras que la encontrada para las posiciones no sinónimas (π_a) es 0.004. En los dos primeros exones del gen la diversidad nucleotídica es 0.01 y bastante similar a la de 0.08 detectada en el tercer exón. En la región no codificadora (regiones 5' y 3' flanqueantes e intrones) la estima de $\pi_{\text{NC}} = 0.012$ es ligeramente inferior a la observada para las posiciones sinónimas ($\pi_s = 0.019$). Las estimas de θ para la mayor parte de las posiciones analizadas son muy parecidas a las estimas de la diversidad nucleotídica.

Tabla IV.1. Estimaciones del polimorfismo nucleotídico del gen CG6296 en *D. melanogaster*

Región	Nº de sitios	S	π	θ
No codificadora	745	28	0.012	0.011
Silenciosa	1237,566	57	0.015	0.013
Codificadora	2025	56	0.008	0.008
Sinónima	492,6	29	0.019	0.017
No sinónima	1532,427	27	0.004	0.005
Exones 1 y 2	317	24	0.01	0.011
Exón 3	1708	43	0.008	0.007
Total	2770	84	0.009	0.008

Test del neutralismo

Para poder establecer si los patrones del polimorfismo observados en el fragmento analizado se ajustan a las predicciones del neutralismo, se aplicaron los test desarrollados por Tajima (1989) y Fu y Li (1993) para diferentes tipos de posiciones (tabla IV.2). Los estadísticos propuestos para los distintos tests fueron negativos tanto para las posiciones no sinónimas como para el total de posiciones de los exones 1 y 2. Sin embargo, los estadísticos son significativos ($P < 0.05$) únicamente para las posiciones no sinónimas de estos exones. Así, los test de Tajima y Fu y Li sugieren que los patrones del polimorfismo detectados en la región CG6296 concuerdan con las predicciones del neutralismo a excepción de las posiciones no sinónimas de los exones 1 y 2 que muestran un exceso de variantes segregando a baja frecuencia.

El test H de Fay y Wu (2000) se aplicó para establecer si existe un exceso de mutaciones derivadas a alta frecuencia (tabla IV.2). *D. simulans* fue utilizada como especie *outgroup* para polarizar los cambios observados. Para todos los tipos de

posiciones analizadas el test H es negativo y es marginalmente significativo para las posiciones no sinónimas de los exones 1 y 2 y para el total de las posiciones codificadoras de estos exones ($P = 0.063$ y 0.062 , respectivamente) bajo la hipótesis conservativa de no recombinación. Sin embargo, el test fue significativo para el total de la región estudiada al considerar un valor de $R_L = 45$ que fue estimado a partir del número mínimo de eventos de recombinación detectado en la muestra (Rozas et al 2001).

Tabla IV.2. Tests del neutralismo en diferentes regiones del gen CG6296 en *D. melanogaster*

Test	Exones 1 y 2					Exón 3			Total Codificadora	Total Región
	No Codificadora	Silenciosas	Sinónima	No Sinónima	Total	Sinónima	No Sinónima	Total		
Tajima D	0.363	0.456	0.235	- 1.861	- 0.543	0.654	0.034	0.257	0.059	0.167
$P(D < D_{obs})$	0.691	0.726	0.642	0.023*	0.306	0.769	0.577	0.7	0.591	0.629
Fu y Li D*	0.475	0.522	0.364	- 2.577	- 0.782	0.697	0.154	0.259	- 0.009	0.164
$P(D < D^*_{obs})$	0.636	0.712	0.519	0.026*	0.254	0.719	0.452	0.567	0.482	0.534
Fu y Li F*	0.514	0.584	0.379	- 2.739	- 0.826	0.794	0.139	0.301	0.013	0.192
$P(F < F^*_{obs})$	0.683	0.752	0.634	0.032*	0.206	0.813	0.364	0.394	0.487	0.581
Fay y Wu	- 5.661	- 9.871	- 2.058	- 1.491	- 3.549	- 2.152	- 2.409	- 4.561	- 8.111	- 13.771
$P(H < H_{obs})$	0.087	0.106	0.115	0.063	0.062	0.161	0.177	0.164	0.12	0.095

Probabilidad de obtener un valor del test estadístico menor que el observado tras simulación de coalescencia sin recombinación.

* $P < 0.05$

El test de MacDonald y Kreitman (1991) se realizó para contrastar si la relación entre el número de polimorfismos dentro de especie y el número de diferencias fijadas entre especies es equivalente para los cambios sinónimos y los no sinónimos. Para aplicar este test se utilizó *D. simulans* en la comparación interespecífica. No se detectaron diferencias significativas entre ambas relaciones (tabla IV.3).

Tabla IV.3 Test de MK (1991) en el gen CG6296 de *D. melanogaster*

	Polimorfismo en <i>D. melanogaster</i>	Diferencias fijadas (<i>D. melanogaster/D. simulans</i>)
No sinónimo	26	41
Sinónimo	27	40
G value = 0.031; $P = 0.859$		
NI = 0.939		

Genealogía de los alelos

La figura IV.4 muestra la genealogía de los alelos de las poblaciones de Montemayor y Sant Sadurn d'Anoia. En este análisis se empleó *D. simulans* como

especie *outgroup*. En el árbol se pueden identificar las 4 líneas que presentan el mismo haplotipo (M2, M11, M22 y M26), las dos que comparte otro haplotipo (CN6 y CN22) y los 13 haplotipos restantes formados por líneas únicas. Además, los alelos no se agrupan por localidad, lo que confirma el hecho de que las poblaciones analizadas no se encuentran genéticamente diferenciadas para la región CG6296.

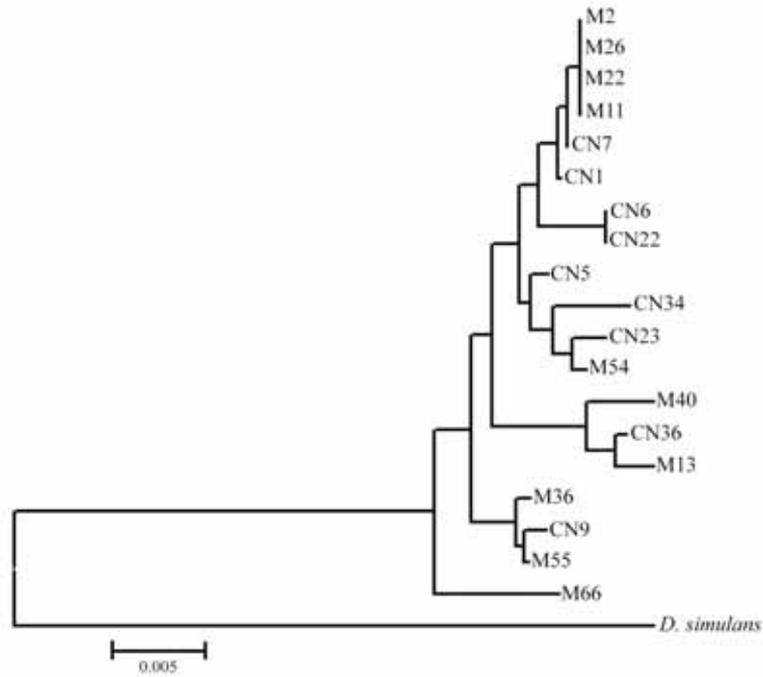


Figura IV.4. Genealogía de los alelos de las poblaciones de Montemayor (M) y Sant Sadurn d'Anoia (CN) reconstruida por el método de *neighbor-joining* (Saitou y Nei 1987) usando la distancia de Kimura de 2 parámetros (Kimura 1980).

IV.1.2 DISCUSIÓN

IV.1.2.1 DIVERSIDAD NUCLEOTÍDICA DEL GEN CG6296 EN *D. MELANOGASTER*

El gen CG6296 de *D. melanogaster* se localiza en el brazo cromosómico 3R (posición 97D) al igual que los genes *Dca* y *Fst*. La variabilidad nucleotídica de los tres genes se ha analizado en las mismas dos poblaciones naturales de *D. melanogaster*: Montemayor y Sant Sadurní d'Anoia. La estima de la diversidad nucleotídica silenciosa detectada en la región CG6296 es idéntica a la observada en la región *Dca* ($\pi_{\text{sil}} = 0.015$) y bastante superior a la correspondiente estima en el gen *Fst* ($\pi_{\text{sil}} = 0.0028$). Según Kliman y Hey (1993), la tasa de recombinación para la región donde se encuentra ubicado el gen *Fst* es de alrededor de 0.00093 mientras que para las regiones *Dca* y CG6296 es muy similar y de aproximadamente 0.00228 y 0.00385, respectivamente. Por lo tanto, el mayor nivel de polimorfismo silencioso detectado en las dos últimas regiones podría ser consecuencia de su ubicación en zonas con mayores niveles en la tasa de recombinación en comparación a los del gen *Fst* en *D. melanogaster*.

Como se ha descrito en otros genes, la variación en las posiciones sinónimas ($\pi_s = 0.019$) del gen CG6296 es superior que la detectada en las regiones 5' ($\pi = 0.014$) y 3' ($\pi = 0.006$) flanqueantes, lo que sugiere una mayor restricción funcional en estas regiones probablemente debido a la presencia de regiones reguladoras. El nivel de variación nucleotídica en la población de Sant Sadurní d'Anoia se ha estimado tras el análisis de 109 fragmentos no codificadores localizados a lo largo del cromosoma X (Orengo y Agudé 2004). El valor promedio de la diversidad nucleotídica en el anterior estudio fue de $\pi = 0.0038$. No obstante, el rango de variación oscila entre 0 y 0.0132 lo que indica que el nivel de variación en la región no codificadora del gen CG6296 de $\pi = 0.012$ correspondería a un nivel de variación elevado en la población. Por otro lado, dentro de la región codificadora, el nivel de la diversidad nucleotídica en los dos primeros exones es bastante similar a la detectada en el tercer exón (tabla IV.1). En el primer exón se encuentra la secuencia del péptido señal de la proteína codificada por el gen CG6296 mientras que el segundo y parte del tercero codifican para un dominio lipasa. De hecho, la posible función descrita para esta

proteína es la de triacilglicerol lipasa que está implicada en el transporte y el metabolismo de lípidos. El nivel de diversidad nucleotídica no sinónima para la región que codifica el dominio lipasa (segundo y parte del tercer exón) es $\pi_a = 0.0025$ y casi tres veces inferior al detectado ($\pi_a = 0.0067$) en el resto de la proteína (resto del tercer exón), indicando una mayor restricción funcional en el dominio lipasa.

Los test de neutralidad de Tajima y Fu y Li (tabla 21) basados en la comparación de la heterocigosidad detectada dentro de las poblaciones analizadas de *D. melanogaster* no fueron significativos (tabla IV.2). Sin embargo, el test H (Fay y Wu 2000) indica que en toda la región analizada del gen CG6296 existe un exceso significativo de mutaciones derivadas segregando a altas frecuencias cuando se emplea el valor estimado de $R_L = 45$ (tabla IV.2). El exceso de mutaciones derivadas a altas frecuencias podría ser causado por un barrido selectivo en o cerca del gen CG6296.

IV.2 DIVERGENCIA DEL GEN CG6296

IV.2.1 RESULTADOS

IV.2.1.1 DIVERGENCIA NUCLEOTÍDICA DEL GEN CG6296 EN EL GRUPO *MELANOGASTER* DE *DROSOPHILA*

El gen CG6296 fue secuenciado en cinco especies del grupo *melanogaster* de *Drosophila*: *D. melanogaster*, *D. simulans*, *D. mauritiana*, *D. yakuba* y *D. erecta*. El alineamiento múltiple de la región codificadora de las cinco especies comprende un total de 2979 posiciones que se reducen a 1848 al excluir de los análisis las posiciones que presentan *gaps*. En todas las especies, la estructura del gen CG6296 esta compuesta por 3 exones separados por dos pequeños intrones. La longitud de la región codificadora incluyendo los intrones es de 2138 pb en *D. melanogaster*, 2214 pb en *D. simulans*, 2747 pb en *D. mauritiana*, 2457 pb en *D. yakuba* y 2479 pb en *D. erecta*.

En la tabla IV.4 se muestran las estimas de la divergencia nucleotídica entre las diferentes especies para el gen CG6296. Las estimas de la divergencia no sinónima (K_a) son más bajas que las detectadas para las posiciones sinónimas (K_s) en todas las comparaciones. Las estimas de la relación entre la divergencia no sinónima y sinónima (K_a/K_s) son inferiores a 1 y van desde 0.210 en la comparación entre *D. simulans* y *D. mauritiana* hasta 0.441 entre *D. yakuba* y *D. erecta*.

Tabla IV. 4. Divergencia nucleotídica en el gen CG6296

Especies	K_s	K_a	K_a/K_s
<i>mel - sim</i>	0.124	0.052	0.419
<i>mel - mau</i>	0.124	0.049	0.395
<i>mel - yak</i>	0.234	0.092	0.393
<i>mel - ere</i>	0.286	0.113	0.395
<i>sim - mau</i>	0.038	0.008	0.210
<i>sim - yak</i>	0.229	0.091	0.397
<i>sim - ere</i>	0.276	0.110	0.398
<i>mau - yak</i>	0.223	0.090	0.403
<i>mau - ere</i>	0.275	0.108	0.392
<i>yak - ere</i>	0.213	0.094	0.441

K_s , divergencia sinónima por posición sinónima. K_a , divergencia no sinónima por posición no sinónima. K_a/K_s , ratio entre la divergencia no sinónima y sinónima.

mel. *D. melanogaster*; *sim*, *D. simulans*; *mau*, *D. mauritiana*; *yak*, *D. yakuba*; *ere*, *D. erecta*.

Los árboles obtenidos por el método de *neighbor-joining* (Saitou y Nei 1987) en base a la divergencia nucleotídica detectada en la región codificadora así como los obtenidos a partir de la divergencia sinónima y no sinónima se presentan en la figura IV.5. La topología de los árboles coincide en todos los casos excepto en la agrupación de *D. yakuba* y *D. erecta* en el árbol reconstruido en base a la divergencia no sinónima. En todos los árboles, la rama del linaje de *D. erecta* es consideradamente más larga que la del linaje de *D. yakuba*.

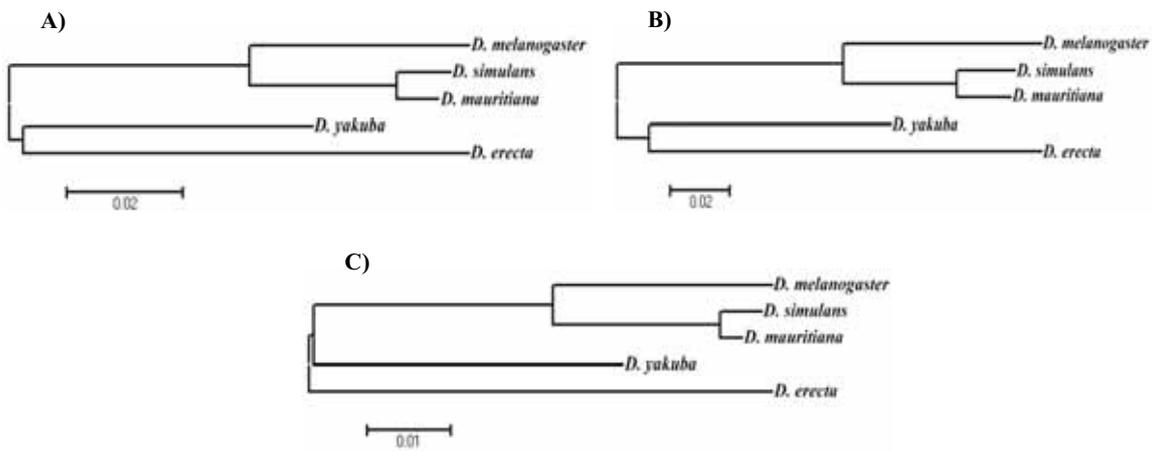


Figura IV.5. Árboles filogenéticos obtenidos por el método de *neighbor-joining* (Saitou y Nei 1987) en base a la A) divergencia nucleotídica total, B) divergencia sinónima y C) divergencia no sinónima.

Para poder determinar si el número de sustituciones entre los linajes de *D. yakuba* y *D. erecta* son significativamente diferentes se aplicó el *relative rate test* propuesto por Wu y Li (1985) utilizando *D. melanogaster* como especie *outgroup* (tabla IV.5). Los resultados del anterior test indican una diferencia estadísticamente significativa en la tasa de sustitución de los linajes de *D. yakuba* y *D. erecta*. Al aplicar el test de Tajima (1993) se pudo comprobar que en el linaje de *D. erecta* se ha producido un exceso significativo tanto de sustituciones sinónimas como no sinónimas (tabla IV.5).

Tabla IV.5 Relative rate test del gen CG6296 en el subgrupo *melanogaster*

Out-Group	Especie 1- Especie 2	Test de Wu y Li			Test de Tajima					
		K ₁₂ - K ₂₁	SD	z - score	Sustituciones sinónimas			Sustituciones no sinónimas		
					m1	m2	χ^2	r1	r2	χ^2
<i>mel</i>	<i>yak - ere</i>	- 0.034446	0.009451	- 3.644533***	38	58	0.041*	45	81	0.001**
<i>sim</i>	<i>yak - ere</i>	- 0.031728	0.009218	- 3.441958***	39	66	0.008**	39	63	0.017*
<i>mau</i>	<i>yak - ere</i>	- 0.042101	0.009801	- 4.295417***	49	95	0.0001***	52	80	0.014*

m1 y m2, número de sustituciones sinónimas en los linajes que llevan a las especies 1 y 2, respectivamente.

r1 y r2, número de sustituciones no sinónimas en los linajes que llevan a las especies 1 y 2, respectivamente.

Nombre de las especies como en la tabla IV.4.

* $P < 0.05$; ** $0.01 < P < 0.05$; *** $P < 0.01$

La estima del sesgo en el uso aleatorio de codones sinónimos fue estimada para las diferentes especies (tabla IV.6). Los niveles de *codon bias* detectados en todas las especies fueron muy similares y relativamente bajos, con una media del valor ENC igual a 54.295. *Drosophila mauritiana* fue la especie con el *codon bias* más alto (ENC = 52.691), mientras que *D. erecta* presenta el más bajo (ENC = 56.246).

Tabla IV.6. Estimaciones de *codon bias* en el gen CG6296

Especies	Scaled χ^2	CBI	ENC
<i>D. melanogaster</i>	0.176	0.250	54.377
<i>D. simulans</i>	0.182	0.259	53.250
<i>D. mauritiana</i>	0.191	0.268	52.691
<i>D. yakuba</i>	0.175	0.258	54.913
<i>D. erecta</i>	0.134	0.216	56.246

CBI, índice de *codon bias*; ENC, número efectivo de codones

IV.2.1.1.1 DIVERGENCIA AMINOACÍDICA DEL GEN CG6296 EN EL GRUPO *MELANOGASTER* DE *DROSOPHILA*

El alineamiento múltiple de la proteína codificada por el gen CG6296 para las 5 especies del subgrupo *melanogaster* comprende un total de 998 aminoácidos (figura IV.6). La longitud de la proteína codificada por el gen CG6296 en *D. melanogaster*, *D. simulans*, *D. mauritiana*, *D. yakuba* y *D. erecta* es de 676, 701, 879, 782 y 862 aminoácidos, respectivamente. Estas variaciones en el tamaño se deben en parte a la variación en el número de repeticiones PEEST perfectos o en aquellos que difieren en

uno de sus residuos (PEETT, PEEAT, PEDST, PEEVT, SEEST, PEETS, PEEDT, PEEET y PEVST) en la región C-terminal de la proteína. Este número oscila entre 7 en *D. yakuba* hasta 15 en *D. mauritiana*. Además de los anteriores motivos, también se detectaron motivos que presentan una o dos inserciones con respecto a los motivos PEEST (PEEDST, PEESST, PEDEST, PEEVST, PEEEST y PEEVEST). Al tener en cuenta estos motivos, la oscilación en el número de motivos abarca desde 10 en *D. yakuba* hasta 18 en *D. mauritiana*.

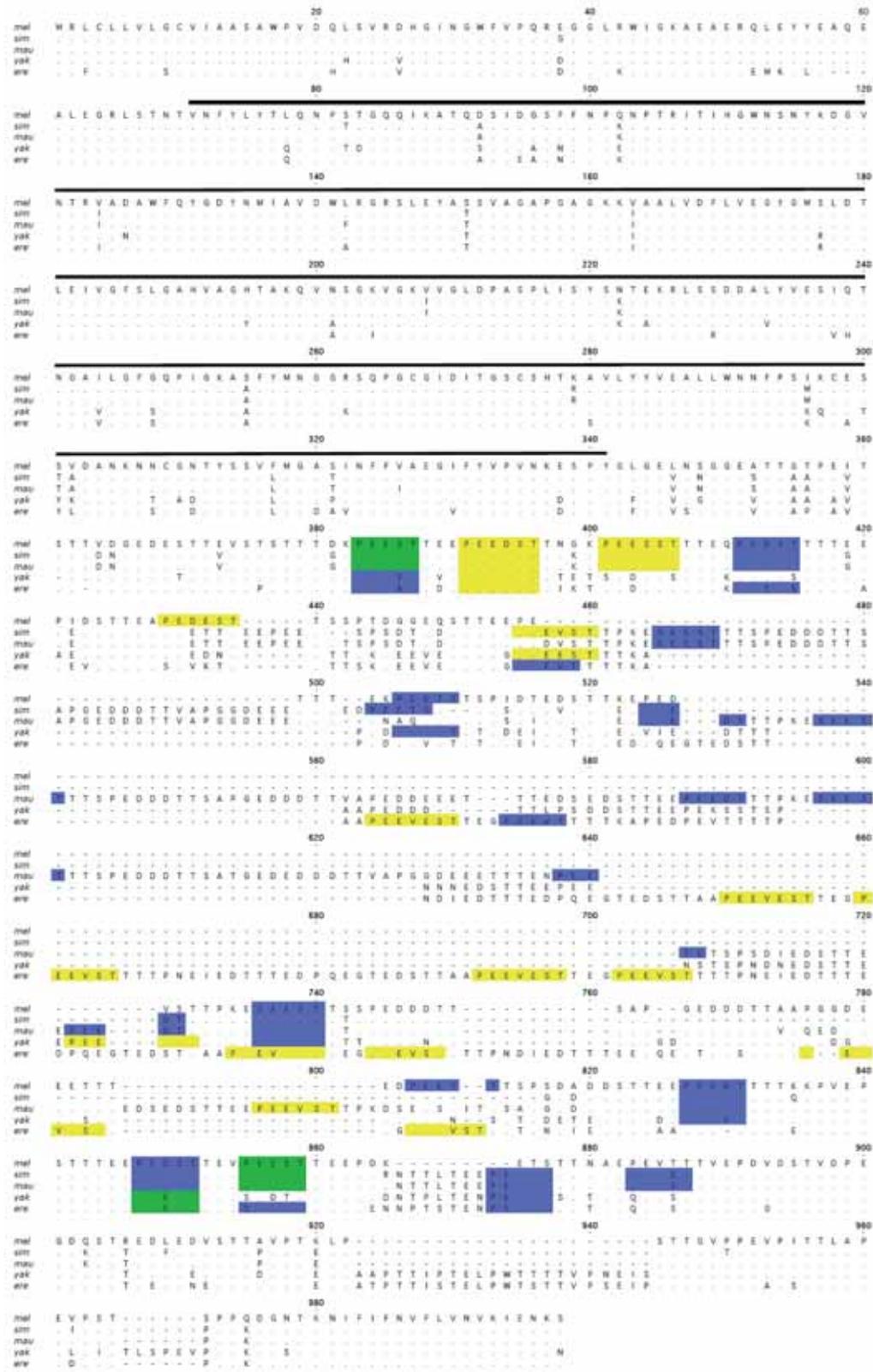


Figura IV.6. Alineamiento múltiple de la secuencia aminoacídica de la proteína codificada por el gen CG6296. Los puntos (.) indican aminoácidos idénticos a los de la secuencia de referencia y los guiones (-) deleciones de aminoácidos respecto a esta secuencia. La línea negra indica el dominio lipasa. En verde se muestran los motivos PEEST; en azul, aquellos motivos que difieren del PEEST en uno de los residuos y en amarillo, los que presentan una inserción de uno o dos residuos con respecto al motivo PEEST. *mel*, *D. melanogaster*; *sim*, *D. simulans*; *mau*, *D. mauritiana*; *yak*, *D. yakuba* y *ere*, *D. erecta*.

IV.2.2 DISCUSIÓN

IV.2.2.1 DIVERGENCIA NUCLEOTÍDICA Y AMINOACÍDICA DEL GEN CG6296 EN EL GRUPO *MELANOGASTER*

La divergencia del gen CG6296 se analizó en 5 especies que se encuentran incluidas en el subgrupo *melanogaster* de *Drosophila*. Entre *D. melanogaster* y *D. simulans*, la estima de divergencia sinónima ($K_s = 0.124$) es superior a la estima promedio de la divergencia nucleotídica silenciosa hallada en 19 genes autosómicos analizados en ambas especies ($K_{sil} = 0.108$; Begun y Whitley 2000). De hecho, únicamente 6 de los 19 genes están evolucionando en las posiciones sinónimas más rápidamente que el gen CG6296. Igualmente, la estima de la divergencia no sinónima de $K_a = 0.052$ es superior a la estima promedio para los mismos genes ($K_a = 0.011$; Begun 2002). En este caso, ninguno de los 19 genes anteriores presentan estimas de K_a superiores a las detectadas en el gen CG6296. Este hecho indicaría que este gen estaría evolucionando bastante rápidamente en las posiciones no sinónimas. No obstante, la relación entre la divergencia no sinónima y sinónima en la comparación entre *D. melanogaster* y *D. simulans* ($K_a/K_s = 0.419$), indicaría que la selección purificadora también actúa sobre el CG6296 (tabla IV.4).

En todos los árboles obtenidos a partir de las diferentes estimas de la divergencia en el gen CG6296, la rama que lleva al linaje de *D. erecta* es inusualmente larga en comparación a la de *D. yakuba*. Empleando *D. melanogaster*, *D. simulans* y *D. mauritiana* como especies *outgroup*, el *relative rate test* detectó un exceso significativo en la tasa de sustitución global en el linaje de *D. erecta* (tabla IV.5). De hecho, la región codificadora del gen CG6296 presenta los niveles más bajos de restricción funcional en la comparación entre *D. yakuba* y *D. erecta* ($K_a/K_s = 0.441$).

Por otro lado, en las especies que conforman el subgrupo *melanogaster* la proteína CG6296 presenta una variación intrarespecífica considerable en el número de motivos PEEST perfectos o en aquellos que difieren en uno de los residuos. Esto indica que el gen CG6296 también está evolucionando de forma rápida por duplicaciones/deleciones internas que contribuyen a importantes alteraciones en su longitud dentro de este grupo. De hecho, la característica más destacable del análisis de la divergencia de la proteína CG6296 es la relativa conservación de aproximadamente los primeros 430 aminoácidos en contraposición con una evolución muy rápida por

adiciones y deleciones en el resto de la proteína. La estructura de esta proteína presenta un dominio lipasa identificado entre los residuos 71 y 341 así como diversos dominios de baja complejidad a partir del residuo 365. El valor de la estima promedio de la relación K_a/K_s en el dominio lipasa de 0.258 contrasta con el observado para el resto de la proteína de 0.395, sugiriendo que es precisamente en las regiones con baja complejidad donde las constricciones funcionales son menores permitiendo detectar una elevada divergencia tanto a nivel aminoacídico como del tamaño de la proteína.

CONCLUSIONES

1- El gen *Fst* presenta un bajo nivel de polimorfismo nucleotídico y un exceso significativo de variantes únicas en *D. melanogaster*. Este resultado indicaría que en o cerca del gen *Fst* se ha producido un fenómeno de arrastre selectivo o *hitchhiking*.

2- El nivel de variación nucleotídica de la región *Fst* en *D. subobscura* no está reducido a pesar de que este gen está localizado muy cerca de uno de los puntos de rotura de la inversión O₇. Este resultado indicaría que esta inversión no es de origen reciente.

3- La región codificadora del gen *Fst* presenta polimorfismo por adiciones y deleciones de longitud considerable tanto en *D. melanogaster* como en *D. subobscura*, aunque únicamente en la primera especie dichas adiciones/deleciones afectan a las repeticiones internas presentes en la mitad terminal de la región codificadora.

4- El patrón de polimorfismo en la región 5' del gen *Dca* no se ajusta a las predicciones del neutralismo ni en *D. melanogaster* ni en *D. subobscura*. Este resultado indicaría la acción de selección en las secuencias reguladoras de la expresión de dicho gen en ambas especies.

5- Las poblaciones de Montemayor y Sant Sadurní d'Anoia de *D. melanogaster* no están genéticamente diferenciadas ni en la región *Fst* ni en la región *Dca* (tampoco en la región CG6296). Este resultado indicaría que las diferencias climáticas entre estas poblaciones no son lo suficientemente drásticas para provocar un efecto en la variabilidad nucleotídica de dichos genes.

6- El linaje de *D. erecta* presenta un exceso significativo de sustituciones no sinónimas en el gen *Fst* que indicaría un acúmulo de cambios adaptativos o una relajación de la selección purificadora en este linaje con respecto al linaje de *D. yakuba*.

7- La proteína FST presenta una región con diversos residuos de prolina consecutivos que se mantiene en todas las especies del subgrupo *melanogaster* pero que está ausente en las especies del grupo *obscura*.

8- El gen *Dca* ha sufrido un proceso de duplicación independiente en los linajes de *D. ananassae*, *D. willistoni* y *D. pseudoobscura* siendo mucho más reciente el producido en este último linaje. Asimismo, se ha detectado una relajación de la selección purificadora en contra de las mutaciones no sinónimas en el linaje de *D. ananassae*, en el que uno de los tres genes parálogos ha sufrido un proceso de pseudogenización.

9- El sesgo en el uso de codones sinónimos del gen *Dca* difiere considerablemente en las especies del subgrupo *melanogaster* en comparación a las del grupo *obscura* y *D. ananassae*, siendo más elevado en estas últimas especies. Este resultado indicaría una relajación de la selección débil en contra de las mutaciones de codón preferente a no preferente en las especies del subgrupo *melanogaster*.

10- El gen *Dca* es parálogo del gen *RC* y ha surgido por una duplicación génica previa a la radiación de las especies del subgénero *Sophophora* pero posterior a la separación de los subgéneros *Sophophora* y *Drosophila*. Tras su origen, el gen *Dca* ha sufrido un proceso de evolución rápida posiblemente ligado a la adquisición de una nueva función.

11- El gen *Dca* muestra una diferencia significativa de la relación $\omega = dn/ds$ en los linajes de las especies de *Drosophila* tropicales y templadas. Este resultado está de acuerdo con la implicación de este gen en la tolerancia al frío.

12- El gen *RC* de *D. melanogaster* presenta un exceso significativo de mutaciones derivadas segregando a elevada frecuencia que podría indicar un fenómeno de arrastre selectivo o *hitchhiking* en o cerca del gen *RC* en esta especie.

13- El gen *RC* presenta un exceso significativo de mutaciones no sinónimas fijadas entre *D. subobscura* y *D. madeirensis* y entre *D. subobscura* y *D. guanche* respecto al polimorfismo en *D. subobscura*. Este resultado indicaría un acúmulo de cambios adaptativos durante la divergencia de las especies del cluster *subobscura* que también se reflejaría en los *relative rate test* realizados utilizando *D. pseudoobscura* como especie de referencia y en la estima de máxima versomilitud de $\omega > 1$ en el linaje de *D. subobscura*.

14- La región terminal con repeticiones internas de las proteínas FST y CG6296 presenta una evolución muy rápida por adiciones/deleciones que provoca importantes variaciones de la longitud de dichas proteínas en las especies analizadas.

BIBLIOGRAFÍA

- AGUADÉ, M., 1999. Positive selection drives the evolution of the Acp29AB accessory gland protein in *Drosophila*. *Genetics*. **152**: 543-551.
- AGUADÉ, M., and C. H. LANGLEY, 1994. Polymorphism and divergence in regions of low recombination in *Drosophila*., pp. 67-76 in *Non-neutral Evolution: Theories and Molecular Data.*, edited by G. B. Chapman and Hall, London.
- AGUADÉ, M., N. MIYASHITA and C. H. LANGLEY, 1989. Reduced Variation in the *yellow-achaete-scute* region in natural populations of *Drosophila melanogaster*. *Genetics*. **122**: 607-615.
- AKASHI, H., 1995. Inferring weak selection from patterns of polymorphism and divergence at "silent" sites in *Drosophila* DNA. *Genetics*. **139**: 1067-1076.
- ANDERSON, A. R., J. E. COLLINGE, A. A. HOFFMANN, M. KELLETT and S. W. MCKECHNIE, 2003. Thermal tolerance trade-offs associated with the right arm of chromosome 3 and marked by the *hsr-omega* gene in *Drosophila melanogaster*. *Heredity*. **90**: 195-202.
- ANDERSON, A. R., A. A. HOFFMANN, S. W. MCKECHNIE, P. A. UMINA and A. R. WEEKS, 2005. The latitudinal cline in the *In(3R)Payne* inversion polymorphism has shifted in the last 20 years in Australian *Drosophila melanogaster* populations. *Mol. Ecol.* **14**: 851-858.
- ANDOLFATTO, P., 2001. Contrasting patterns of X-linked and autosomal nucleotide variation in *Drosophila melanogaster* and *Drosophila simulans*. *Mol. Biol. Evol.* **18**: 279-290.
- ANDOLFATTO, P., and M. PRZEWORSKI, 2000. A genome-wide departure from the standard neutral model in natural populations of *Drosophila*. *Genetics*. **156**: 257-268.
- ANDOLFATTO, P. and PRZEWORSKI, M. 2001. Regions of lower crossing over harbor more rare variants in African populations of *Drosophila melanogaster*. *Genetics*. **158**: 657-665.
- AQUADRO, C. F., D. J. BEGUN and E. C. KINDAHL, 1994. Selection, recombination, and DNA polymorphism in *Drosophila*, pp. 46-56 in *Non-Neutral Evolution: Theories and Molecular Data*, edited by B. GOLDING. Chapman and Hall, New York.
- ASHBURNER, M., 1989. *Drosophila: A laboratory handbook*. Cold Spring Harbor Laboratory Press, New York.
- AYRINHAC A, DEBAT V, GILBERT P, KISTER AG, LEGOUT H, MORETEAU B, 2004. Cold adaptation in geographical populations of *Drosophila melanogaster*: phenotypic plasticity is more important than genetic variability. *Funct. Ecol.* **18**: 700–706.

- BEGUN, D. J., 2002. Protein variation in *Drosophila simulans*, and comparison of genes from centromeric versus noncentromeric regions of chromosome 3. *Mol. Biol. Evol.* **19**: 201-203.
- BEGUN, D. J., and C. F. AQUADRO, 1992. Levels of naturally occurring DNA polymorphism correlate with recombination rates in *D. melanogaster*. *Nature*. **356**: 519-520.
- BEGUN, D. J., and P. WHITLEY, 2000a. Reduced X-linked nucleotide polymorphism in *Drosophila simulans*. *Proc. Natl. Acad. Sci. USA*. **97**: 5960-5965.
- BEGUN, D. J., and P. WHITLEY, 2000b. Adaptive evolution of relish, a *Drosophila* NF-kappaB/IkappaB protein. *Genetics*. **154**: 1231-1238.
- BERRY, A. J., J. W. AJIOKA and M. KREITMAN, 1991. Lack of polymorphism on the *Drosophila* fourth chromosome resulting from selection. *Genetics*. **129**: 1111-1117.
- BETTENCOURT, B. R., I. KIM, A. A. HOFFMANN and M. E. FEDER, 2002. Response to natural and laboratory selection at the *Drosophila hsp70* genes. *Evolution. Int. J. Org. Evolution*. **56**: 1796-1801.
- BIERNE, N., and A. EYRE-WALKER, 2004. The genomic rate of adaptive amino acid substitution in *Drosophila*. *Mol. Biol. Evol.* **21**: 1350-1360.
- BIRKY, C. W., JR., and J. B. WALSH, 1988. Effects of linkage on rates of molecular evolution. *Proc Natl. Acad. Sci. USA*. **85**: 6414-6418.
- BRIDGES, C. B. 1935. Salivary chromosome maps. *J. Hered.* **26**:60-64
- BUBLI, O.A., IMASHEVA, A.G. and LOESCHCKE, V., 1998. Selection for knockdown resistance to heat in *Drosophila melanogaster* at high and low larval densities. *Evolution*. **52**, pp. 619–625.
- CAPY, P., PLA, E. and DAVID, J.R., 1993. Phenotypic and genetic variability of morphometrical traits in natural populations of *Drosophila melanogaster* and *D. simulans*. 1. Geographic variations. *Genet. Select. Evol.* **25**, pp. 517–536
- CAVICCHI, S., GUERRA, D., LA TORRE, V. and HUEY, R.B., 1995. Chromosomal analysis of heat-shock tolerance in *Drosophila melanogaster* evolving at different temperatures in the laboratory. *Evolution*. **49**, pp. 676–684.
- CHAO, H., F. D. SONNICHSEN, C. I. DELUCA, B. D. SYKES and P. L. DAVIES, 1994. Structure-function relationship in the globular type III antifreeze protein: identification of a cluster of surface residues required for binding to ice. *Protein Sci.* **3**: 1760-1769.
- CHARLESWORTH, B., 1994. Genetic recombination. Patterns in the genome. *Curr. Biol.* **4**: 182-184.

- CIRERA, S., and M. AGUADÉ, 1997. Evolutionary history of the sex-peptide (Acp70A) gene region in *Drosophila melanogaster*. *Genetics*. **147**: 189-197.
- CIRERA, S., and M. AGUADÉ, 1998. Molecular evolution of a duplication: the sex-peptide (Acp70A) gene region of *Drosophila subobscura* and *Drosophila madeirensis*. *Mol. Biol. Evol.* **15**: 988-996.
- CIVETTA, A. and R.S. SINGH, 1995. High divergence of reproductive tract proteins and their association with postzygotic reproductive isolation in *Drosophila melanogaster* and *Drosophila virilis* group species. *J. Mol. Biol. Evol.* **41**: 1085-1095.
- CLARK, A. G., 1994. Invasion and maintenance of a gene duplication. *Proc. Natl. Acad. Sci. USA.* **91**: 2950-2954.
- COMERON, J. M. 1994. Estudi de la variabilitat nucleotídica a *Drosophila*: Regió *Xdh* a *D. subobscura*. Tesi doctoral. Universitat de Barcelona, Barcelona.
- COMERON, J. M., and M. KREITMAN, 2000. The correlation between intron length and recombination in *Drosophila*. Dynamic equilibrium between mutational and selective forces. *Genetics*. **156**: 1175-1190.
- COOKE, J., M. A. NOWAK, M. BOERLIJST and J. MAYNARD-SMITH, 1997. Evolutionary origins and maintenance of redundant gene expression during metazoan development. *Trends. Genet.* **13**: 360-364.
- COSSINS, A. R., M. BEHAN, G. JONES and K. BOWLER, 1987. Lipid-protein interactions in the adaptive regulation of membrane function. *Biochem. Soc. Trans.* **15**: 77-81.
- DAHLGAARD, J., LOESCHCKE, V., MICHALAK, P. and JUSTESEN, J., 1998. Induced thermotolerance and associated expression of the heat-shock protein Hsp70 in adult *Drosophila melanogaster*. *Funct. Ecol.* **12**, pp. 786-793.
- DAVID, J. R., and P. CAPY, 1988. Genetic variation of *Drosophila melanogaster* natural populations. *Trends. Genet.* **4**: 106-111.
- DO, C. B., M. S. MAHABHASHYAM, M. BRUDNO and S. BATZOGLOU, 2005. ProbCons: Probabilistic consistency-based multiple sequence alignment. *Genome Res.***15**: 330-340.
- DROSOPHILA* 12 GENOMES CONSORTIUM. 2007. *Nature.* **450**: 203-218.
- DUMAN, J. G., N. LI, D. VERLEYE, F. W. GOETZ, D. W. WU, 1998. Molecular characterization and sequencing of antifreeze proteins from larvae of the beetle *Dendroides canadensis*. *J. Comp. Physiol. [B]*. **168**: 225-232.
- FAY, J. C., and C. I. WU, 2000. Hitchhiking under positive Darwinian selection. *Genetics*. **155**: 1405-1413.

FAY, J. C., G. J. WYCKOFF and C. I. WU, 2002. Testing the neutral theory of molecular evolution with genomic data from *Drosophila*. *Nature*. **415**: 1024-1026.

FEDER, M. E., N. V. CARTANO, L. MILOS, R. A. KREBS and S. L. LINDQUIST, 1996. Effect of engineering Hsp70 copy number on Hsp70 expression and tolerance of ecologically relevant heat shock in larvae and pupae of *Drosophila melanogaster*. *J. Exp. Biol.* **199**: 1837-1844.

FEDER, M. E., and R. A. KREBS, 1997. Ecological and evolutionary physiology of heat shock proteins and the stress response in *Drosophila*: complementary insights from genetic engineering and natural variation. *Exs.* **83**: 155-173.

FORCE, A., M. LYNCH, F. B. PICKETT, A. AMORES, Y. L. YAN and J. POSTLETHWAIT. 1999. Preservation of duplicate genes by complementary, degenerative mutations. *Genetics* **151**: 1531-1545.

FU, Y. X., and W. H. LI, 1993. Statistical tests of neutrality of mutations. *Genetics*. **133**: 693-709.

FUJITA, T., 1999. Senescence marker protein-30 (SMP30): structure and biological function. *Biochem. Biophys. Res. Commun.* **254**: 1-4.

FUJITA, T., H. INOUE, T. KITAMURA, N. SATO, T. SHIMOSAWA and N. MARUYAMA. 1998. Senescence marker protein-30 (SMP30) rescues cell death by enhancing plasma membrane Ca(2+)-pumping activity in Hep G2 cells. *Biochem. Biophys. Res. Commun.* **250**: 374-380.

FUJITA, T., J. L. MANDEL, T. SHIRASAWA, O. HINO, T. SHIRAI and N. MARUYAMA. 1995. Isolation of cDNA clone encoding human homologue of senescence marker protein-30 (SMP30) and its location on the X chromosome. *Biochim. Biophys. Acta.* **1263**: 249-252.

FUJITA, T., and N. MARUYAMA, 1991. Elevated levels of c-jun and c-fos transcripts in the aged rat liver. *Biochem. Biophys. Res. Commun.* **178**: 1485-1491.

FUJITA, T., T. SHIRASAWA, K. UCHIDA and N. MARUYAMA, 1992. Isolation of cDNA clone encoding rat senescence marker protein-30 (SMP30) and its tissue distribution. *Biochim. Biophys. Acta.* **1132**: 297-305.

FUJITA, T., T. SHIRASAWA, K. UCHIDA and N. MARUYAMA, 1996. Gene regulation of senescence marker protein-30 (SMP30): coordinated up-regulation with tissue maturation and gradual down-regulation with aging. *Mech. Ageing. Dev.* **87**: 219-229.

FUJITA, T., K. UCHIDA and N. MARUYAMA, 1992. Purification of senescence marker protein-30 (SMP30) and its androgen-independent decrease with age in the rat liver. *Biochim. Biophys. Acta.* **1116**: 122-128.

- GIBERT, P., and R. B. HUEY, 2001. Chill-coma temperature in *Drosophila*: effects of developmental temperature, latitude, and phylogeny. *Physiol. Biochem. Zool.* **74**: 429-434.
- GILAD, Y., S. ROSENBERG, M. PRZEWORSKI, D. LANCET and K. SKORECKI, 2002. Evidence for positive selection and population structure at the human MAO-A gene. *Proc. Natl. Acad. Sci. USA* .**99**: 862-867.
- GLINKA, S., OMETTO, L., MOUSSET, S., STEPHAN, W., and DE LORENZO, D. 2003. Demography and natural selection have shaped genetic variation in *Drosophila melanogaster*: A multi-locus approach. *Genetics* **165**: 1269-1278.
- GODDARD, K., A. CACCONE, and J.R. POWELL, 1990. Evolutionary implications of DNA divergence in the *Drosophila obscura* group. *Evolution*. **44**: 1656-1670.
- GOTO, S. G., 2000. Expression of *Drosophila* homologue of senescence marker protein-30 during cold acclimation. *J. Insect. Physiol.* **46**: 1111-1120.
- GOTO, S. G., 2001. A novel gene that is up-regulated during recovery from cold shock in *Drosophila melanogaster*. *Gene*. **270**: 259-264.
- GRAHAM, L. A., Y. C. LIOU, V. K. WALKER and P. L. DAVIES, 1997. Hyperactive antifreeze protein from beetles. *Nature*. **388**: 727-728.
- GREENBERG, A. J., J. R. MORAN, J. A. COYNE and C. I. WU, 2003. Ecological adaptation during incipient speciation revealed by precise gene replacement. *Science*. **302**: 1754-1757.
- GU, J., Y. WANG and X. GU, 2002. Evolutionary analysis for functional divergence of Jak protein kinase domains and tissue-specific genes. *J. Mol. Evol.* **54**: 725-733.
- GUERRA, D., CAVICCHI, S., KREBS, R.A. AND LOESCHCKE, V., 1997. Resistance to heat and cold stress in *Drosophila melanogaster*—intra and inter population variation in relation to climate. *Genet. Select. Evol.* **29**, pp. 497–510.
- HANSEN, I. A., S. R. MEYER, I. SCHAFER and K. SCHELLER, 2002. Interaction of the anterior fat body protein with the hexamerin receptor in the blowfly *Calliphora vicina*. *Eur. J. Biochem.* **269**: 954-960.
- HARR, B., KAUER, M., and SCHLÖTTERER, C. 2002. Hitchhiking mapping: A population-based fine-mapping strategy for adaptive mutations in *Drosophila melanogaster*. *Proc. Natl. Acad. Sci.* **99**: 12949-12954.
- HOFFMANN, A.A., ANDERSON, A. and HALLAS, R., 2002. Opposing clines for high and low temperature resistance in *Drosophila melanogaster*, *Ecol. Lett.* **5**, pp. 614–618.
- HOFFMANN AA, SORENSEN JG, LOESCHCKE V., 2003. Adaptation of *Drosophila* to temperature extremes: bringing together quantitative and molecular approaches. *J. Thermal. Biol.* **28**: 175–216.

- HUDSON, R. R., 1983. Properties of a neutral allele model with intragenic recombination. *Theor. Popul. Biol.* **23**: 183-201.
- HUDSON, R. R., 1987. Estimating the recombination parameter of a finite population model without selection. *Genet. Res.* **50**: 245-250.
- HUDSON, R. R., 1990. Gene genealogies and the coalescent process, pp. 1-44 in *Oxford Surveys in Evolutionary Biology*, edited by J. ANTONOVICS and D. FUTUYMA. Oxford University Press, Oxford.
- HUDSON, R. R., and N. L. KAPLAN, 1985. Statistical properties of the number of recombination events in the history of a sample of DNA sequences. *Genetics.* **111**: 147-164.
- HUDSON, R. R., M. KREITMAN and M. AGUADE, 1987. A test of neutral molecular evolution based on nucleotide data. *Genetics.* **116**: 153-159.
- HUDSON, R. R., M. SLATKIN and W. P. MADDISON. 1992. Estimation of levels of gene flow from DNA sequence data. *Genetics.* **132**: 583-589.
- HUGHES, A. L., 1999. Genomic catastrophism and the origin of vertebrate immunity. *Arch. Immunol. Ther. Exp. (Warsz)* **47**: 347-353.
- INGRAM, V. M. 1961. Gene evolution and the haemoglobins. *Nature.* **189**:704-708.
- INGRAM, V. M. 1963. *The hemoglobins in genetics and evolution.* Columbia University Press, New York.
- KAUER, M., ZANGERL, B., DIERINGER, D., and SCHLÖTTERER, C. 2002. Chromosomal patterns of microsatellite variability contrast sharply in African and Non-African populations of *Drosophila melanogaster*. *Genetics.* **160**: 247-256.
- KAUER, M.O., DIERINGER, D., and SCHLÖTTERER, C. 2003. A microsatellite variability screen for positive selection associated with the "out of Africa" habitat expansion of *Drosophila melanogaster*. *Genetics* **165**: 1137-1148.
- KARAN, D., A. K. MUNJAL, P. GIBERT, B. MORETEAU, R. PARKASH and J.R. DAVID. 1998. Latitudinal clines for morphometrical traits in *Drosophila kikkawai*: a study of natural populations from the Indian subcontinent. *Genet. Res.* **71**: 31-38.
- KELLY, J. K., 1997. A test of neutrality based on interlocus associations. *Genetics.* **146**: 1197-1206.
- KIMURA, M., 1968. Evolutionary rate at the molecular level. *Nature.* **217**: 624-626.
- KIMURA, M., 1979. The neutral theory of molecular evolution. *Sci. Am.* **241**: 98-100, 102, 108 passim.
- KIMURA, M., 1983. *The Neutral Theory of Molecular Evolution.* Cambridge University Press, Cambridge.

- KIMURA, M. T., 2004. Cold and heat tolerance of drosophilid flies with reference to their latitudinal distributions. *Oecologia*. **140**: 442-449.
- KINGMAN, J. F. C., 1982. On the genealogy of large populations. *J. Appl. Probab.* **19A**: 27-43.
- KLEINEIDAM, R. G., P. A. JEKEL, J. J. BEINTEMA and P. SITUMORANG, 1999. Seminal-type ribonuclease genes in ruminants, sequence conservation without protein expression? *Gene*. **231**: 147-153.
- KLIMAN, R. M., and J. HEY, 1993. Reduced natural selection associated with low recombination in *Drosophila melanogaster*. *Mol. Biol. Evol.* **10**: 1239-1258.
- KNIGHT, C. A., C. C. CHENG and A. L. DEVRIES, 1991. Adsorption of alpha-helical antifreeze peptides on specific ice crystal surface planes. *Biophys. J.* **59**: 409-418.
- KO, W. Y., R. M. DAVID and H. AKASHI, 2003. Molecular phylogeny of the *Drosophila melanogaster* species subgroup. *J. Mol. Evol.* **57**: 562-573.
- KREBS, R. A., 1999. A comparison of Hsp70 expression and thermotolerance in adults and larvae of three *Drosophila* species. *Cell Stress Chaperones*. **4**: 243-249.
- KREBS, R. A., and M. E. FEDER, 1997. Deleterious consequences of *Hsp70* overexpression in *Drosophila melanogaster* larvae. *Cell Stress Chaperones*. **2**: 60-71.
- KREBS, R. A., and M. E. FEDER, 1998. Hsp70 and larval thermotolerance in *Drosophila melanogaster*: how much is enough and when is more too much? *J. Insect. Physiol.* **44**: 1091-1101.
- KREBS, R. A., V. LA TORRE, V. LOESCHCKE and S. CAVICCHI, 1996. Heat-shock resistance in *Drosophila* populations: analysis of variation in reciprocal cross progeny. *Hereditas*. **124**: 47-55.
- KREBS, R. A., and V. LOESCHCKE, 1996. Acclimation and selection for increased resistance to thermal stress in *Drosophila buzzatii*. *Genetics*. **142**: 471-479.
- KUMAR, S., K. TAMURA and M. NEI, 2004. MEGA3: Integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment. *Brief. Bioinform.* **5**: 150-163.
- LAAYOUNI, H., F. GARCÍA-FRANCO, B. E. CHÁVEZ-SANDOVAL, V. TROTTA, S. BELTRAN, M. COROMINAS and M. SANTOS. 2007. Thermal evolution of genes expression profiles in *Drosophila subobscura*. *BMC Evol Biol.* **7**: 42.
- LANGE, B. W., C. H. LANGLEY and W. STEPHAN, 1990. Molecular evolution of *Drosophila* metallothionein genes. *Genetics*. **126**: 921-932.
- LERMAN, D. N., and M. E. FEDER, 2001. Laboratory selection at different temperatures modifies heat-shock transcription factor (HSF) activation in *Drosophila melanogaster*. *J. Exp. Biol.* **204**: 315-323.

LEWIS, E. B. 1951. Pseudoallelism and gene evolution. Cold Spring Harbor Symp. Quant. Biol. **16**:159-174.

LEWONTIN, R.C. 1974. The interactions of selection and linkage. I. General considerations: heterotic models. Genetics. **49**: 49-67.

LI, W. H., 1997. Molecular Evolution. Sinauer Press, Sunderland, MA.

LIU, Y. C., P. THIBAUT, V. K. WALKER, P. L. DAVIES and L. A. GRAHAM, 1999. A complex family of highly heterogeneous and internally repetitive hyperactive antifreeze proteins from the beetle *Tenebrio molitor*. Biochemistry. **38**: 11415-11424.

LIU, Y. C., A. TOCILJ, P. L. DAVIES and Z. JIA, 2000. Mimicry of ice structure by surface hydroxyls and water of a beta-helix antifreeze protein. Nature. **406**: 322-324.

LLOPART, A. 1999. Evolución de la región *RpII215* en el grupo *obscura* de *Drosophila* y comparación con el grupo *melanogaster* : Análisis a nivel nucleotídico . Tesi doctoral. Universitat de Barcelona, Barcelona.

LLOPART, A., and M. AGUADÉ. 1999. Synonymous rates at the *RpII215* gene of *Drosophila*: variation among species and across the coding region. Genetics. **152**: 269-280.

LOESCHCKE, V. and KREBS, R.A., 1996. Selection for heat-shock resistance in larval and in adult *Drosophila buzzatii*: comparing direct and indirect responses. Evolution. **50**, pp. 2354–2359.

LONG, M., E. BETRAN, K. THORNTON and W. WANG, 2003. The origin of new genes: glimpses from the young and old. Nat. Rev. Genet. **4**: 865-875.

LYNCH, M., and J. S. CONERY, 2000. The evolutionary fate and consequences of duplicate genes. Science. **290**: 1151-1155.

LYNCH, M., and A. FORCE, 2000. The probability of duplicate gene preservation by subfunctionalization. Genetics. **154**: 459-473.

MADDISON, W. P., and D. R. MADDISON, 1989. Interactive analysis of phylogeny and character evolution using the computer program MacClade. Folia Primatol (Basel) **53**: 190-202.

MARTIN-CAMPOS, J. M., J. M. COMERON, N. MIYASHITA and M. AGUADÉ, 1992. Intraspecific and interspecific variation at the *y-ac-sc* region of *Drosophila simulans* and *Drosophila melanogaster*. Genetics. **130**: 805-816.

MARTIN-CAMPOS, J. M. 1998. Análisis de la variabilidad a nivel nucleotídico en la región *6-Pgd* en *D. subobscura*. Tesi doctoral. Universitat de Barcelona, Barcelona.

MAYNARD SMITH, J., and J. HAIGH, 1974. The hitch-hiking effect of a favourable gene. Genet. Res. **23**: 23-35.

- McCOLL, G., A. A. HOFFMANN and S. W. MCKECHNIE, 1996. Response of two heat shock genes to selection for knockdown heat resistance in *Drosophila melanogaster*. *Genetics*. **143**: 1615-1627.
- McCOLL, G., and S. W. MCKECHNIE, 1999. The *Drosophila* heat shock *hsr-omega* gene: an allele frequency cline detected by quantitative PCR. *Mol. Biol. Evol.* **16**: 1568-1574.
- McDONALD, J. H., and M. KREITMAN, 1991. Adaptive protein evolution at the Adh locus in *Drosophila*. *Nature*. **351**: 652-654.
- MONTGOMERY, E., B. CHARLESWORTH and 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**: 31-41.
- MORGAN, T. J., and T. F. MACKAY, 2006. Quantitative trait loci for thermotolerance phenotypes in *Drosophila melanogaster*. *Heredity*. **96**: 232-242.
- MORIYAMA, E. N., and J. R. POWELL, 1996. Intraspecific nuclear DNA variation in *Drosophila*. *Mol. Biol. Evol.* **13**: 261-277.
- MORTON, B. R., 1993. Chloroplast DNA codon use: evidence for selection at the psb A locus based on tRNA availability. *J. Mol. Evol.* **37**: 273-280.
- MOUNT, S. M., C. BURKS, G. HERTZ, G. D. STORMO, O. WHITE and C. FIELDS. 1992. Splicing signals in *Drosophila*: intron size, information content, and consensus sequences. *Nucleic. Acids. Res.* **20**: 4255-4262.
- MOUSSET, S., BRAZIER, L., CARIOU, M.L., CHARTOIS, F., DEPAULIS, F., and VEUILLE, M. 2003. Evidence of a high rate of selective sweeps in African *Drosophila melanogaster*. *Genetics*. **163**: 599-609.
- MULLER, H. J. 1936. Bar duplication. *Science*. **83**:528-530.
- MUNTE, A., M. AGUADÉ and C. SEGARRA, 1997. Divergence of the *yellow* gene between *Drosophila melanogaster* and *D. subobscura*: recombination rate, codon bias and synonymous substitutions. *Genetics*. **147**: 165-175.
- MUNTE, A., M. AGUADÉ and C. SEGARRA, 2000. Nucleotide variation at the *yellow* gene region is not reduced in *Drosophila subobscura*: a study in relation to chromosomal polymorphism. *Mol. Biol. Evol.* **17**: 1942-1955.
- MUNTE, A., M. AGUADÉ and C. SEGARRA, 2001. Changes in the recombinational environment affect divergence in the *yellow* gene of *Drosophila*. *Mol. Biol. Evol.* **18**: 1045-1056.
- MUNTE, A., J. ROZAS, M. AGUADÉ and C. SEGARRA, 2005 Chromosomal inversion polymorphism leads to extensive genetic structure: a multilocus survey in *Drosophila subobscura*. *Genetics*. **169**: 1573-1581.

NACHMAN, M. W., S. N. BOYER and C. F. AQUADRO, 1994. Nonneutral evolution at the mitochondrial NADH dehydrogenase subunit 3 gene in mice. *Proc. Natl. Acad. Sci. U S A.* **91**: 6364-6368.

NACHMAN, M. W., W. M. BROWN, M. STONEKING and C. F. AQUADRO, 1996. Nonneutral mitochondrial DNA variation in humans and chimpanzees. *Genetics.* **142**: 953-963.

NAKAJIMA, Y., and S. NATORI, 2000. Identification and characterization of an anterior fat body protein in an insect. *J. Biochem. (Tokyo)* **127**: 901-908.

NAVARRO-SABATE, A., M. AGUADÉ and C. SEGARRA, 1999 The relationship between allozyme and chromosomal polymorphism inferred from nucleotide variation at the *AcpH-1* gene region of *Drosophila subobscura*. *Genetics.* **153**: 871-889.

NEI, M., 1987. *Molecular Evolutionary Genetics*. Columbia University Press, New York.

NEI, M., and T. GOJOBORI, 1986. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol. Biol. Evol.* **3**: 418-426.

NEI, M., and S. KUMAR, 2000. *Molecular evolution and phylogenetics*. Oxford University Press, Oxford.

NEI, M., I. B. ROGOZIN and H. PIONTKIVSKA, 2000. Purifying selection and birth-and-death evolution in the ubiquitin gene family. *Proc. Natl. Acad. Sci. USA.* **97**: 10866-10871.

NIELSEN, R., and Z. YANG, 1998. Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. *Genetics.* **148**: 929-936.

NORRY, F. M., J. DAHLGAARD and V. LOESCHCKE, 2004. Quantitative trait loci affecting knockdown resistance to high temperature in *Drosophila melanogaster*. *Mol. Ecol.* **13**: 3585-3594.

NÓBREGA., C. 2006. A multilocus comparative study of nucleotide variation in *Drosophila madeirensis* and the close relative *D. subobscura*: insights into the speciation process of *D. madeirensis*. Tese de doutoramento. Universidade da Madeira. Portugal.

O'GRADY, P. M., and M. G. KIDWELL, 2002. Phylogeny of the subgenus *sophophora* (Diptera: drosophilidae) based on combined analysis of nuclear and mitochondrial sequences. *Mol. Phylogenet. Evol.* **22**: 442-453.

OHNO, S., 1970. *Evolution by gene duplication*. Springer, Berlin.

OHTA, T., 1992. The nearly neutral theory of molecular evolution. *Annu. Rev. Ecol. Syst.* **23**: 263-286.

- OHTA, T., 1994. Further examples of evolution by gene duplication revealed through DNA sequence comparisons. *Genetics*. **138**: 1331-1337.
- ORENGO, D. J., and M. AGUADÉ, 2004. Detecting the footprint of positive selection in a european population of *Drosophila melanogaster*: multilocus pattern of variation and distance to coding regions. *Genetics*. **167**: 1759-1766.
- ORENGO, D. J., and M. AGUADÉ, 2007. Genome Scans of Variation and Adaptive Change: Extended Analysis of a Candidate Locus Close to the *phantom* Gene Region in *Drosophila melanogaster*. *Mol. Biol. Evol.* **24**: 1122-1129.
- PARSCH, J., 2003. Selective constraints on intron evolution in *Drosophila*. *Genetics*. **165**: 1843-1851.
- PENALVA, O. F., H. SAKAMOTO, A. NAVARRO-SABATÉ, E. SAKASHITA, B. GRANADINO, C. SEGARRA and L. SÁNCHEZ. 1996. Regulation of the gene *Sex-lethal*: A comparative analysis of *Drosophila melanogaster* and *Drosophila subobscura*. *Genetics*. **144**: 1653-1664.
- PEROTTI, M., F. TODDEI, F. MIRABELLI, M. VAIRETTI, G. BELLOMO. D.J. McCONKEY and S. ORRENIUS. 1990. Calcium-dependent DNA fragmentation in human synovial cells exposed to cold shock. *FEBS Lett.* **259**: 331-334.
- PIONTKIVSKA, H., A. P. ROONEY and M. NEI, 2002. Purifying selection and birth-and-death evolution in the histone H4 gene family. *Mol. Biol. Evol.* **19**: 689-697.
- POWELL, J. R., 1997. Progress and prospects in evolutionary biology: the *Drosophila* model. Oxford University Press, New York.
- PRITCHARD, J. K., and M. PRZEWORSKI, 2001. Linkage disequilibrium in humans: models and data. *Am. J. Hum. Genet.* **69**: 1-14.
- PRZEWORSKI, M., J. D. WALL and P. ANDOLFATTO, 2001. Recombination and the frequency spectrum in *Drosophila melanogaster* and *Drosophila simulans*. *Mol. Biol. Evol.* **18**: 291-298.
- QIN, W., S. J. NEAL, R. M. ROBERTSON, J. T. WESTWOOD and V. K. WALKER, 2005. Cold hardening and transcriptional change in *Drosophila melanogaster*. *Insect. Mol. Biol.* **14**: 607-613.
- RAMOS-ONSINS, S., and M. AGUADÉ, 1998. Molecular evolution of the Cecropin multigene family in *Drosophila*. functional genes vs. pseudogenes. *Genetics*. **150**: 157-171.
- RAND, D. M., and L. M. KANN, 1996. Excess amino acid polymorphism in mitochondrial DNA: contrasts among genes from *Drosophila*, mice, and humans. *Mol. Biol. Evol.* **13**: 735-748.
- ROZAS, J., C. SEGARRA, C. ZAPATA, G. ALVAREZ, and M. AGUADÉ. 1995. Nucleotide polymorphism at the *rp49* region of *Drosophila subobscura*: Lack of

geographic subdivision within chromosomal arrangements in Europe. *J. Evol. Biol.* **8**: 355-367.

ROZAS, J., J. C. SÁNCHEZ-DELBARRIO, X. MESSEGUER and R. ROZAS, 2003. DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics* **19**: 2496-2497.

ROZAS, J., and R. ROZAS, 1999. DnaSP version 3: an integrated program for molecular population genetics and molecular evolution analysis. *Bioinformatics.* **15**: 174-175.

ROZAS, J., M. GULLAUD, G. BLANDIN and M. AGUADÉ, 2001. DNA Variation at the *rp49* Gene Region of *Drosophila simulans*: Evolutionary Inferences From an Unusual Haplotype Structure. *Genetics.* **158**: 1147-1155.

SAITOU, N., and M. NEI, 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* **4**: 406-425.

SANCHEZ-GRACIA, A., 2005. Evolución molecular de los genes del sistema olfatorio *OS-E* y *OS-F* en diferentes especies de *Drosophila*. Tesis doctoral. Universitat de Barcelona, Barcelona.

SANCHEZ-GRACIA, A., M. AGUADÉ and J. ROZAS, 2003. Patterns of nucleotide polymorphism and divergence in the Odorant-Binding Protein genes *OS-E* and *OS-F*: Analysis in the melanogaster species subgroup of *Drosophila*. *Genetics.* **165**: 1279-1288.

SARICH, V. M., and A. C. WILSON, 1973. Generation time and genomic evolution in primates. *Science.* **179**: 1144-1147.

SEGARRA, C., and M. AGUADÉ, 1992. Molecular organization of the X chromosome in different species of the obscura group of *Drosophila*. *Genetics.* **130**: 513-521.

SHI, L., M. SAWADA, U. SESTER and J. C. CARLSON, 1994. Alterations in free radical activity in aging *Drosophila*. *Exp. Gerontol.* **29**: 575-584.

SHIELDS, D.C., P.M. SHARP, D.G. HIGGINS, and F. WRIGHT, 1988. "Silent" sites in *Drosophila* genes are not neutral: evidence of selection among synonymous codons. *Mol. Biol. Evol.* **5**: 704-716.

SHIMOKAWA, N., and M. YAMAGUCHI, 1993. Molecular cloning and sequencing of the cDNA coding for a calcium-binding protein regucalcin from rat liver. *FEBS Lett.* **327**: 251-255.

SINCLAIR, B. J., A. ADDO-BEDIAKO and S. L. CHOWN, 2003. Climatic variability and the evolution of insect freeze tolerance. *Biol. Rev. Camb. Philos. Soc.* **78**: 181-195.

SINCLAIR, B. J., A. G. GIBBS and S. P. ROBERTS, 2007. Gene transcription during exposure to, and recovery from, cold and desiccation stress in *Drosophila melanogaster*. *Insect. Mol. Biol.* **16** (4): 435-443.

SINGH, N. D., P. F. ARNDT and D. A. PETROV. 2006. Minor shift in background substitutional patterns in the *Drosophila saltans* and *willistoni* lineages is insufficient to explain GC content of coding sequences. *BMC Biol.* **4**: 37.

SMITH, N. G., and A. EYRE-WALKER, 2002. Adaptive protein evolution in *Drosophila*. *Nature.* **415**: 1022-1024.

STEPHAN, W., 1994. Effects of genetic recombination and population subdivision on nucleotide sequence variation in *Drosophila annanasae.*, pp. 57-66. in *Non-neutral Evolution: Theories and Molecular Data.*, edited by G. B. Chapman and Hall, London.

STEPHAN, W., V. S. RODRIGUEZ, B. ZHOU and J. PARSCH, 1994. Molecular evolution of the metallothionein gene *Mtn* in the melanogaster species group: results from *Drosophila ananassae*. *Genetics.* **138**: 135-143.

STEPHENS, S. G. 1951. Possible significance of duplication in evolution. *Adv. Genet.* **4**:247-265.

STOREY, K. B., 1999. Living in the cold: freeze-induced gene responses in freeze-tolerant vertebrates. *Clin. Exp. Pharmacol. Physiol.* **26**: 57-63.

STOREY, K. B., and J. M. STOREY, 1986. Freeze tolerance and intolerance as strategies of winter survival in terrestrially-hibernating amphibians. *Comp. Biochem. Physiol. A.* **83**: 613-617.

STOREY, K. B., and J. M. STOREY, 1988. Freeze tolerance in animals. *Physiol. Rev.* **68**: 27-84.

SWANSON, W. J., C. F. AQUADRO and V. D. VACQUIER, 2001. Polymorphism in abalone fertilization proteins is consistent with the neutral evolution of the egg's receptor for lysin (VERL) and positive darwinian selection of sperm lysin. *Mol. Biol. Evol.* **18**: 376-383.

TAJIMA, F., 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics.* **123**: 585-595.

TAJIMA, F., 1993. Simple methods for testing the molecular evolutionary clock hypothesis. *Genetics.* **135**: 599-607.

TAKAHASHI, H., and M. YAMAGUCHI, 1993. Regulatory effect of regucalcin on (Ca²⁺)-Mg²⁺)-ATPase in rat liver plasma membranes: comparison with the activation by Mn²⁺ and Co²⁺. *Mol. Cell. Biochem.* **124**: 169-174.

TAKAHASHI, H., and M. YAMAGUCHI, 1993. Regucalcin modulates hormonal effect on (Ca²⁺)-Mg²⁺)-ATPase activity in rat liver plasma membranes. *Mol. Cell. Biochem.* **125**: 171-177.

TAKAHASHI, H., and M. YAMAGUCHI, 1994. Activating effect of regucalcin on (Ca²⁺)-Mg²⁺)-ATPase in rat liver plasma membranes: relation to sulfhydryl group. *Mol. Cell. Biochem.* **136**: 71-76.

TAVARE, S., D. J. BALDING, R. C. GRIFFITHS and P. DONNELLY, 1997. Inferring coalescence times from DNA sequence data. *Genetics*. **145**: 505-518.

THOMPSON, J. D., T. J. GIBSON, F. PLEWNIAK, F. JEANMOUGIN and D. G. HIGGINS, 1997. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic. Acids. Res.* **25**: 4876-4882.

TRABESINGER-RUEF, N., T. JERMANN, T. ZANKEL, B. DURRANT, G. FRANK and S.A. BENNER. 1996. Pseudogenes in ribonuclease evolution: a source of new biomacromolecular function? *FEBS Lett.* **382**: 319-322.

TSAUR, S. C., and C. I. WU, 1997. Positive selection and the molecular evolution of a gene of male reproduction, *Acp26Aa* of *Drosophila*. *Mol Biol Evol* **14**: 544-549.

VICARIO, S., E. N. MORIYAMA and J. R. POWELL. 2007. Codon usage in twelve species of *Drosophila*. *BMC Evol Biol.* **7**: 226.

WAGNER, A., 2000. Decoupled evolution of coding region and mRNA expression patterns after gene duplication: implications for the neutralist-selectionist debate. *Proc. Natl. Acad. Sci. USA.* **97**: 6579-6584.

WAKELEY, J., and N. ALIACAR, 2001. Gene genealogies in a metapopulation. *Genetics*. **159**: 893-905.

WALL, J. D., 1999. Recombination and the power of statistical tests of neutrality. *Genet. Res.* **73**: 65-79.

WALL, J.D., ANDOLFATTO, P., and PRZEWORSKI, M. 2002. Testing models of selection and demography in *Drosophila simulans*. *Genetics* **162**: 203-216.

WALSH, J. B., 1995. How often do duplicated genes evolve new functions? *Genetics*. **139**: 421-428.

WATTERSON, G. A., 1975. On the number of segregating sites in genetical models without recombination. *Theor. Popul. Biol.* **7**: 256-276.

WHELAN, S., and N. GOLDMAN, 1999. Distributions of Statistics Used for the Comparison of Models of Sequence Evolution in Phylogenetics. *Mol. Biol. Evol.* **16**: 1292-1299.

WRIGHT, F., 1990. The 'effective number of codons' used in a gene. *Gene*. **87**: 23-29.

WU, C. I., and W. H. LI, 1985. Evidence for higher rates of nucleotide substitution in rodents than in man. *Proc. Natl. Acad. Sci. USA.* **82**: 1741-1745.

YAMAGUCHI, M., 2005. Role of regucalcin in maintaining cell homeostasis and function (review). *Int. J. Mol. Med.* **15**: 371-389.

- YAMAGUCHI, M., and T. YAMAMOTO, 1978. Purification of calcium binding substance from soluble fraction of normal rat liver. *Chem. Pharm. Bull. (Tokyo)* **26**: 1915-1918.
- YANG, Z., 1997. PAML: a program package for phylogenetic analysis by maximum likelihood. *Comput. Appl. Biosci.* **13**: 555-556.
- YANG, Z., 1998. Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution. *Mol. Biol. Evol.* **15**: 568-573.
- YANG, Z., 2002. Inference of selection from multiple species alignments. *Curr. Opin. Genet. Dev.* **12**: 688-694.
- YANG, Z., and J.P. BIELAWSKI, 2000. Statistical methods for detecting molecular adaptation. *Trends. Ecol. Evol.* **15**: 496-503.
- YANG, Z., and R. NIELSEN, 2000. Estimating synonymous and nonsynonymous substitution rates under realistic evolutionary models. *Mol. Biol. Evol.* **17**: 32-43.
- YANG, Z., and R. NIELSEN, 2002. Codon-substitution models for detecting molecular adaptation at individual sites along specific lineages. *Mol. Biol. Evol.* **19**: 908-917.
- YANG, Z., R. NIELSEN, N. GOLDMAN and A. M. PEDERSEN, 2000. Codon-substitution models for heterogeneous selection pressure at amino acid sites. *Genetics*. **155**: 431-449.
- YANG, Z., and W. J. SWANSON, 2002. Codon-substitution models to detect adaptive evolution that account for heterogeneous selective pressures among site classes. *Mol. Biol. Evol.* **19**: 49-57.
- YU, J., Z. YANG, M. KIBUKAWA, M. PADDOCK, D. A. PASSEY and G. KA-SHU WONG. 2002. Minimal introns are not "junk". *Genome. Res.* **12**: 1185-1189.
- ZAPATA, C., and G. ALVAREZ, 1993. On the detection of nonrandom associations between DNA polymorphisms in natural populations of *Drosophila*. *Mol. Biol. Evol.* **10**: 823-841.
- ZHANG, J., R. NIELSEN and Z. YANG, 2005. Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level. *Mol. Biol. Evol.* **22**: 2472-2479.