

Universitat de Barcelona

Facultat de Biologia
Departament de Genètica

Evolución molecular
de los genes del sistema olfatorio *OS-E* y *OS-F*
en diferentes especies de *Drosophila*

Alejandro Sánchez-Gracia

Barcelona, Noviembre de 2005

4. Resumen de los resultados y discusión general

La contribución relativa de la selección natural positiva (selección Darwiniana) a la evolución del DNA es una de las cuestiones más antiguas y debatidas en evolución molecular. Durante más de medio siglo la biología evolutiva ha estado dominada por la visión Darwiniana de que los organismos se adaptan progresivamente al medio mediante la fijación de mutaciones beneficiosas. De la misma forma, los evolucionistas esperaban que este principio también se extendiera a nivel molecular. No obstante, el cambio revolucionario que generó la aplicación de las técnicas moleculares en el estudio del polimorfismo genético, en los años 1960, resultó en el inicio de una gran controversia acerca del mecanismo de mantenimiento de la variabilidad genética en las poblaciones naturales (KIMURA and OHTA 1971; LEWONTIN 1974; NEI 1975; AYALA 1976). Particularmente importante fue el impacto de la teoría neutralista de la evolución molecular, postulada por KIMURA en el año 1968 (KIMURA 1968), que propone que una gran proporción de la variación molecular existente dentro de las poblaciones sería neutra, y que la mayor parte de las sustituciones nucleotídicas se producirían por deriva genética. Esta teoría puede ser usada como hipótesis nula para detectar las desviaciones de este modelo producidas por la acción de la selección natural positiva, negativa o por procesos de tipo demográfico.

El estudio de la adaptación a nivel molecular es un tema que ha suscitado un enorme interés por su gran impacto en la biología evolutiva. Una de las estrategias utilizadas para detectar la acción de la selección natural en las secuencias de DNA es la de analizar la variabilidad en genes candidatos, es decir en genes que por su función son susceptibles de haber incorporado frecuentemente cambios de tipo adaptativo y que, por lo tanto, tengan una alta probabilidad de que se pueda detectar la huella dejada por la selección natural en el nivel y patrón de variabilidad.

En este sentido, los genes del sistema olfativo son *-a priori-* unos buenos candidatos ya que forman parte de un sistema que participa en la interacción del individuo con el medio externo, siendo crítico en funciones biológicas esenciales como la obtención de comida, la defensa ante predadores, la reproducción y el comportamiento social. De hecho, se ha demostrado que la selección natural positiva ha promovido la fijación de mutaciones beneficiosas durante la evolución de estos genes en diferentes

especies (NGAI *et al.* 1993; WILLET 2000; KRIEGER and ROSS 2002; GILAD *et al.* 2003; CLARK *et al.* 2003; EMES *et al.* 2004; WATTS *et al.* 2004).

En *Drosophila*, los genes *OS-E* y *OS-F* codifican para proteínas de unión a odorantes (OBP) y constituyen, por lo tanto, uno de los primeros contactos del individuo con el exterior. Esta posición tan externa en la cascada de genes que transmiten la señal olfatoria proporciona una cierta flexibilidad evolutiva, que por ende puede posibilitar que sean diana de la selección natural positiva. De hecho, los genes de la familia de las OBP tienen diferentes especificidades de unión, se expresan en diferentes localizaciones citológicas y presentan una elevada diversidad de secuencia; todo ello sugiere que han evolucionado diversificándose funcionalmente dentro de una misma actividad general. Tanto en insectos como en vertebrados, los genes de las OBPs constituyen una familia multigénica con un gran número de miembros. Desde una óptica evolutiva, el análisis de los genes *OS-E* y *OS-F* es, asimismo, interesante en el estudio de la evolución de las duplicaciones génicas. Así, el análisis de la variabilidad de estos genes en diferentes especies puede aportar una importante información sobre los mecanismos de generación y preservación de las copias duplicadas en el genoma, del papel de la selección natural en estos procesos, y su efecto global en la evolución molecular de las grandes familias multigénicas, y por ende del propio genoma.

4.1 Polimorfismo y divergencia nucleotídica de la región OS en el grupo melanogaster de *Drosophila*

Se ha analizado la variabilidad nucleotídica intraespecífica de la región que incluye los genes *OS-E* y *OS-F* en poblaciones naturales de dos especies del subgénero Sophophora: *Drosophila melanogaster* y *D. simulans*. Los niveles de polimorfismo silencioso (incluyendo tanto el polimorfismo nucleotídico en posiciones sinónimas, como en posiciones no codificadoras) estimados como diversidad nucleotídica, π_{SIL} , son diferentes en las dos especies, siendo siempre mayores en *D. simulans*. La población menos polimórfica es la europea (Montemayor) de *D. melanogaster* ($\pi_{\text{SIL}} = 0.0018$), mientras que la más variable es la población Africana de *D. simulans* ($\pi_{\text{SIL}} = 0.020$). Estas diferencias en el nivel de polimorfismo entre ambas especies podrían haber sido causadas por i) diferencias en el tamaño efectivo poblacional de las poblaciones/especies, ii) efectos locales de la región OS en alguna de las especies, iii) por algún evento demográfico ocurrido en la historia evolutiva de estas poblaciones y cuyos efectos sobre la variabilidad aun perdurarían, o bien iv) por el efecto de alguna inversión cromosómica.

4.1.1. Polimorfismo nucleotídico en *D. melanogaster* y divergencia con otras especies del grupo melanogaster

Aunque existe alguna inversión cromosómica fijada entre *D. melanogaster* y *D. simulans*, ninguna afecta a la región donde se localizan los genes *OS-E* y *OS-F*, de manera que un cambio en el ambiente recombinacional inducido por la inversión cromosómica no puede ser el responsable de la diferencia en el nivel de polimorfismo silencioso entre las dos especies. Por otro lado, existen evidencias que sugieren que *D. melanogaster* tiene un tamaño efectivo poblacional menor que *D. simulans* (AKASHI 1996; MOUSSET and DEROME 2004). Estas diferencias en el tamaño efectivo podrían justificar el menor nivel de variación intraespecífica observado en la región OS en *D. melanogaster*. No obstante, la diversidad nucleotídica silenciosa media estimada en varios loci de *D. melanogaster* es siempre mayor que la de la

región OS (MORIYAMA and POWELL 1996, ANDOLFATTO 2001, GLINKA *et al.* 2003; ORENGO and AGUADÉ 2004; KERN and BEGUN 2005), por lo que en la región OS debe de existir algún factor añadido responsable de una reducción local de la variabilidad silenciosa. Asimismo, no parece que diferencias en la tasa de mutación neutra entre las especies puedan justificar las diferencias observadas en el patrón del polimorfismo; de hecho, el nivel de divergencia silenciosa de la región OS es muy similar al estimado en otras regiones entre estas dos mismas especies (e.g. ORENGO and AGUADÉ 2004).

La región OS de *D. melanogaster* se localiza en la banda cromosómica 83CD, cerca del centrómero del brazo cromosómico 3R, en una zona donde existen evidencias de que la recombinación está reducida (KLIMAN and HEY 1993; TRUE *et al.* 1996). En estas zonas de baja recombinación, el efecto de arrastre causado por las mutaciones selectivamente favorables (arrastre selectivo; MAYNARD SMITH and HAIGH 1974) o las mutaciones deletéreas (selección de fondo, CHARLESWORTH *et al.* 1993) en su camino hacia la fijación o la pérdida será muy importante (debido al mayor efecto de ligamento) provocando una reducción significativa de la variabilidad. Este efecto no debería afectar significativamente a los niveles de divergencia interespecífica (no se espera una reducción del número de sustituciones entre las especies para tiempos de divergencia no excesivamente pequeños). El test de *HKA* (HUDSON *et al.* 1987), que permite comprobar estas desviaciones, indica que los niveles de polimorfismo son significativamente menores de lo esperado, y que por lo tanto, el arrastre selectivo podría ser el principal responsable de la reducción de la variabilidad observada en la región OS de *D. melanogaster*.

Una de las maneras de distinguir el efecto de la selección natural positiva, del de la selección de fondo, es mediante el estudio del espectro de frecuencias de las mutaciones. La fijación de mutaciones seleccionadas favorablemente produce un barrido selectivo que elimina toda la variación nucleotídica ligada, de manera que las nuevas mutaciones aparecerán sobre un fondo sin variabilidad. Por lo tanto, durante el periodo posterior al arrastre selectivo, y antes de recuperarse por completo la variabilidad inicial (como media $4N_e$ generaciones), habrá un exceso de variantes segregando a baja frecuencia (genealogía en estrella). Como la aparición mutaciones deletéreas y la posterior eliminación de los cromosomas que las contienen se produce de

forma aleatoria por toda la población, en este caso no se observará una desviación significativa en el espectro de frecuencias de las mutaciones. El efecto de la selección de fondo podría ser descrito, por lo tanto, por el de una disminución del tamaño efectivo en una genealogía neutra. El espectro de frecuencias de las mutaciones que segregan en la región OS de *D. melanogaster*, lejos de presentar un desplazamiento hacia variantes raras, muestra un exceso significativo de mutaciones a frecuencias intermedias (capítulo 3.1, figura 1 y tabla 3). Esta observación es difícil de reconciliar con las hipótesis selectivas descritas anteriormente. Además, la reducción de los niveles de polimorfismo se produce únicamente en la región del gen OS-F: las estimas de variabilidad silenciosa para el gen OS-E y la región intergénica son mucho mayores (capítulo 3.1, tabla 1), y dentro del rango de las estimas medias de *D. melanogaster*. Los valores positivos (y significativos) de los tests basados en la frecuencia de mutaciones también se dan en esta región, donde además se observa una estructuración genética significativamente mayor de la esperada bajo el modelo neutro (capítulo 3.1, tabla 2). Para comprobar si esta estructuración se extendía hacia la región 5' del gen OS-E, o bien se localizaba únicamente sobre este gen, se amplió la región de estudio en unas 2kb en dirección 5'. El patrón observado en esta nueva región es mucho más parecido al de la región del gen OS-F, reduciéndose de nuevo tanto la diversidad nucleotídica silenciosa ($\pi_{\text{SIL}} = 0.002$), como la estructuración de la variabilidad (capítulo 3.1, figura 4). Estos resultados indican que existe un nivel de polimorfismo reducido, en relación al esperado en esta especie, en la mayor parte de la región OS, y que posiblemente esté causado por procesos de tipo selectivo sobre variantes ligadas en una región de recombinación reducida.

Alternativamente, el patrón de variabilidad observado en la región del gen OS-E se podría explicar por algún otro tipo de proceso local, como por ejemplo, por acción de la selección balanceadora sobre alguna, o algunas, posiciones nucleotídicas cercanas. Existen diferentes mecanismos que pueden explicar el polimorfismo balanceado, como la ventaja del heterocigoto, o la selección dependiente de frecuencias. Este tipo de selección puede incrementar tanto la diversidad nucleotídica como la estructuración genética alrededor de la mutación favorable. En el gen OS-E, no obstante, no se ha

encontrado ninguna mutación no sinónima segregando en la población de Montemayor, lo que parecería descartar que exista algún tipo de selección balanceadora que actuara a nivel proteico. Existe la posibilidad, de todas maneras, de que alguna de las mutaciones silenciosas identificadas en la región pueda ser la verdadera responsable, bien actuando a nivel del mRNA (RNA mensajero), o bien a nivel de regulación génica. Aunque los resultados del test *HKA* en esta región no indican que exista un incremento significativo de la variabilidad silenciosa en relación a la divergencia, el nivel basal de variabilidad podría estar ya reducido -tal como se ha comentado anteriormente- debido a los efectos de arrastre en la región que comprende los genes *OS-E* y *OS-F* en esta especie. En estas circunstancias el test no tendría potencia suficiente para detectar el incremento de variabilidad provocado por la acción de la selección balanceadora sobre la región.

Otra hipótesis que podría explicar el patrón de variabilidad nucleotídica observado en la región *OS* de *D. melanogaster* sería la presencia simultánea de varias mutaciones selectivamente favorables que interferirían sus dinámicas de fijación, generando zonas estructuradas (modelo de tráfico, KIRBY and STEPHAN 1996). Según este modelo, la región del gen *OS-E* estaría siendo afectada por varios arrastres selectivos localizados en regiones más o menos cercanas. Es difícil comprobar esta hipótesis; los niveles de variabilidad son muy bajos y sería necesario obtener información más detallada de las regiones cercanas y estudiar con más profundidad los posibles puntos de selección y su distribución en la región. No obstante, en un análisis multilocus del cromosoma X, y en una población también europea, ORENGO and AGUADÉ (2004) obtuvieron valores negativos del estadístico *D* de Tajima que correlacionaban positivamente con la proximidad a las a las zonas codificadoras (posibles dianas para la selección positiva); esta observación apoyaría la existencia de múltiples mutaciones ventajosas que coincidirían en su camino a la fijación en las poblaciones europeas de *D. melanogaster*.

Un aspecto destacado es el patrón atípico de los cambios no sinónimos en el gen *OS-E*. Mientras que no se ha detectado ninguna mutación no sinónima segregando en las 14 líneas de la población Cordobesa de *D. melanogaster*, se han identificado 6 cambios de aminoácido en el linaje que lleva a esta especie. En el gen *OS-F*, en cambio, sólo se han detectado dos

sustituciones no sinónimas fijadas. De hecho, usando la información de la secuencia del gen *OS-E* de *D. erecta* para polarizar los cambios, el número de reemplazamientos fijados en el linaje de *D. melanogaster* es casi significativo ($P = 0.06$). Este exceso de sustituciones se detecta también al analizar completamente toda la zona codificadora (RRT; $P = 0.02$). El test de McDONALD and KREITMAN (1991) indica que no existe un comportamiento diferencial de las mutaciones sinónimas y no sinónimas entre polimorfismo y divergencia; de todas maneras, debido a que sólo existe un polimorfismo en la zona codificadora del gen *OS-E* el test es bastante conservativo. Al añadir los datos del polimorfismo nucleotídico de *D. simulans* (ver sección siguiente), si que se detecta un desacoplamiento de variantes sinónimas y no sinónimas entre polimorfismo y divergencia ($P = 0.005$). Así, aunque se detecta una aceleración global en el linaje de *D. melanogaster* (tanto en el número de sustituciones sinónimas como no sinónimas), parece mucho más acentuada en el caso de las sustituciones que provocan cambios de amino ácido.

Este exceso de sustituciones no sinónimas podría estar provocado por la fijación de 1) mutaciones selectivamente favorables, o de 2) mutaciones ligeramente deletéreas $|N_e s| \approx 1$ debido a variaciones en el tamaño efectivo de la población y/o la reducción de la eficacia de la selección natural (debida al efecto de ligamiento) (OHTA and KIMURA 1971; OHTA 1972). En un escenario de relajación de la selección, también esperaríamos un incremento del número de polimorfismos no sinónimos a bajas frecuencias. Este no es el caso del gen *OS-E* aunque, como hemos comentado anteriormente, el reducido número de mutaciones detectado en la región no permite testar esta predicción. Como las mutaciones sinónimas hacia codones no preferentes pueden comportarse como ligeramente deletéreas, también se esperaría un incremento de este tipo de sustituciones en el linaje de *D. melanogaster*. Del total de sustituciones sinónimas informativas entre *D. melanogaster* y *D. simulans*, sólo 2 son hacia codones no preferentes y 1 hacia preferente. Nuevamente, el reducido número de datos no permite hacer ningún tipo de inferencia sobre una posible relajación. De todas maneras, tanto si se trata de selección direccional positiva, como de una relajación de la selección purificadora, estas fuerzas tendrían que haber actuado de forma diferencial sobre los genes *OS-E* y *OS-F*. Por lo tanto, o bien se han fijado un mayor número de mutaciones selectivamente

ventajosas en el gen *OS-E*, o bien presenta unos niveles de constricción funcional menores que el *OS-F*.

Finalmente, a pesar de que tanto el análisis de variabilidad intraespecífica silenciosa, como el estudio de la divergencia en posiciones codificadoras indican un comportamiento atípico dentro o alrededor de la región del gen *OS-E*, no parece que estos resultados estén relacionados, o por lo menos, no existe una única hipótesis que pueda explicar satisfactoriamente todas las observaciones.

4.1.2. Polimorfismo nucleotídico en poblaciones naturales de *D. simulans*

En *D. simulans* se han secuenciado 22 líneas provenientes de dos poblaciones naturales, 11 de una población europea (Montblanc, España) y otras 11 de una población africana (Maputo, Mozambique). El análisis de la variabilidad nucleotídica indica que las dos poblaciones están genéticamente diferenciadas y que las estimas de variabilidad silenciosa en la población europea ($\pi_{\text{SIL}} = 0.009$) son menores que en la población africana ($\pi_{\text{SIL}} = 0.020$). La distribución de la variabilidad silenciosa a lo largo de la región *OS* en esta especie presenta un patrón algo diferente al observado en *D. melanogaster*. No se observa un gradiente de diversidad nucleotídica a lo largo de la región, aunque se confirma la presencia de una región altamente conservadas dentro del primer intrón del gen *OS-F*. Al contrario que en *D. melanogaster*, no se detecta una desviación significativa de la relación esperada bajo el modelo neutro entre polimorfismo y divergencia. Las estimas, obtenidas a partir de los datos, de la tasa de recombinación en la región *OS* de *D. simulans* son bastante mayores que en la región homóloga de *D. melanogaster*, lo que concuerda con un mayor efecto del ligamento en *D. melanogaster* y podría explicar los resultados obtenidos en la región *OS* de *D. simulans*. Cuando aplicamos el test *MK* (utilizando tanto las 22 secuencias conjuntamente, como en las dos poblaciones por separado), usando *D. melanogaster* para las estimas de divergencia, el resultado es altamente significativo, cosa que no ocurre usando *D. mauritiana* o *D. erecta*, lo que vuelve a apuntar al linaje de *D. melanogaster* como responsable último de la desviación. A pesar de que el

número de polimorfismos nucleotídicos en *D. simulans* es mayor que en *D. melanogaster*, no se detecta ninguno que implique un cambio de amino ácido; además, todos los cambios de amino ácido fijados entre las dos especies, excepto uno, ocurren en el linaje de *D. melanogaster*. Por lo tanto, la hipótesis más probable que explicaría los resultados significativos del test *MK* (ver sección anterior) sería la de un exceso de sustituciones no sinónimas en el linaje de *D. melanogaster*.

La población de Montblanc de *D. simulans* destaca por la alta y atípica estructuración de la variabilidad nucleotídica. Esta estructuración no se encuentra en la muestra africana y es debida a la presencia de un grupo de líneas (haplogrupo #1) formado por secuencias idénticas (o casi idénticas; sólo se diferencian por una única mutación) segregando a frecuencias intermedias (capítulo 3.2, figura 2). El resto de secuencias de esta población (haplogrupo #2) presenta mucha más variabilidad y se encuentra más cercano al equilibrio. Las simulaciones de coalescencia (usando el modelo neutro como hipótesis nula) demuestran, incluso para el caso extremadamente conservativo de recombinación cero, que existe un defecto significativo de diversidad haplotípica, de número de haplotipos y de líneas idénticas, y un exceso de desequilibrio de ligamento y estructuración en la muestra (capítulo 3.2, tabla 4).

La diferencia en los niveles de variabilidad nucleotídica entre poblaciones europeas y africanas de *D. simulans* ya se había observado en otros estudios (HAMBLIN and VEUILLE 1999; ANDOLFATTO 2001). *D. simulans*, al igual que *D. melanogaster*, es una especie de origen africano que se extendió hacia zonas templadas coincidiendo con el desarrollo de la agricultura (hace unos 3000-10000 años; DAVID and CAPY 1988; LACHAISE *et al.* 1988). El menor nivel de variabilidad de la población derivada (europea) podría ser debido a un efecto de cuello de botella ocurrido en la salida de África durante el proceso de colonización de nuevos ambientes. Simultáneamente a dicha colonización pudieron ocurrir otros procesos de tipo adaptativo, ya sea por la fijación de nuevas mutaciones ventajosas en los nuevos ambientes, o bien por la adquisición de ventaja selectiva de mutaciones ya existentes.

En algunos estudios se ha observado que las poblaciones de *D. simulans*, tanto europeas como africanas, presentan una cierta estructuración de la variabilidad. En ningún caso, no obstante, se había descrito un patrón de

estructuración tan extenso como el observado en la región OS. Existen varios escenarios demográficos que podrían generar una estructuración de la variación similar a la observada en *D. simulans*. Por un lado, HAMBLIN and VEUILLE (1999) sugirieron que las poblaciones derivadas de esta especie se habrían formado recientemente por una mezcla de poblaciones ancestrales (africanas) genéticamente diferenciadas. Por otro lado, ANDOLFATTO and PZREWORSKY (2000), ANDOLFATTO (2001) y WALL *et al.* (2002) demostraron que aunque algún aspecto de los datos podría ser compatible con un cuello de botella poblacional, ningún modelo evolutivo simple, demográfico o selectivo, explicaría por sí mismo el patrón de variabilidad de esta especie. Debido a que las dos poblaciones estudiadas están genéticamente diferenciadas, siendo la población europea mucho menos variable que la africana, parecería que el escenario más plausible que explicaría los datos sería el de un cuello de botella poblacional durante la colonización de Europa. No obstante, bajo este escenario no se esperaría una estructuración genética con un grupo de secuencias idénticas (o casi idénticas), y otro grupo mucho más variable y más cercano al equilibrio. Esta configuración parece más probable en un modelo de mezcla de dos poblaciones altamente diferenciadas, una de ellas con un tamaño poblacional muy pequeño. No conocemos si esta situación pudo ocurrir en el origen de la población de Montblanc, aunque no es la explicación más sencilla.

El conjunto de secuencias idénticas podría reflejar también el incremento en la frecuencia de una mutación beneficiosa situada en -o cerca- de dicha región (e.g. un arrastre selectivo, completo o parcial). Este escenario ha sido propuesto por ROZAS *et al.* (2001) para explicar la similar estructuración haplotípica detectada en la región genómica que incluye el gen *rp49*; en este caso, sin embargo, la estructuración se detectaba tanto en una población africana como en una europea (las mismas poblaciones que las estudiadas en este trabajo). En un análisis posterior de la población africana se demostró que la estructura decaía gradualmente con la distancia, tal como se esperaría para un arrastre selectivo en una región de recombinación alta (QUESADA *et al.* 2003). En la región OS la estructuración haplotípica inusual está presente únicamente en la población europea, lo que podría reflejar, quizás, algún proceso de adaptación de tipo local ocurrido en la población europea. Algunos

resultados obtenidos en diferentes poblaciones de *D. simulans* usando datos de microsatélites son consistentes con este escenario (SCHOFEL and SCHLOTTERER 2004). De todas maneras, no se puede descartar por completo que el origen del arrastre selectivo se haya originado en la población africana. De hecho, el alelo mayoritario (haplogrupo #1), que segrega a frecuencias intermedias en Europa, se encuentra también en la población africana, aunque a frecuencias mucho más bajas (se detectó en una única línea). Además, la estimación del tiempo del origen del arrastre selectivo (~10000 años) es compatible con esta última hipótesis. No obstante, para determinar claramente el papel de estas fuerzas evolutivas en la formación del patrón de variabilidad observado en la región OS de *D. simulans* haría falta explorar otros modelos evolutivos, tanto selectivos como demográficos, utilizando información empírica de más poblaciones (sobre todo africanas).

4.2 Polimorfismo nucleotídico y cromosómico en *D. subobscura* y divergencia dentro del grupo *obscura* de *Drosophila*.

Otra de los grupos del subgénero *Sophophora* donde se ha analizado el polimorfismo nucleotídico es el grupo *obscura*, concretamente en la especie *D. subobscura*. La localización cromosómica de la región OS se determinó mediante hibridación *in situ* sobre cromosomas politénicos (utilizando la secuencia de la región OS de *D. melanogaster* como sonda). Los genes OS-E y OS-F se localizaron en la banda cromosómica 98D, en el cromosoma O de *D. subobscura*. Una vez secuenciada y caracterizada la región, se analizó la variabilidad nucleotídica en varias líneas de dos poblaciones naturales, 14 líneas de una población de Galicia y 15 líneas de una población de Bizerta, Túnez. Estas líneas presentan un rico polimorfismo cromosómico por inversiones que afecta a la región OS y que determina que la muestra presente dos clases diferentes de ordenaciones cromosómicas, 1) aquellas que incluyen la región OS -la inversión O_{23^-} y que provienen de la población de Túnez (ordenación O_{3+4+23}) y 2) las que no la incluyen que provienen de la población Gallega ($O_{[3+4]}$) (capítulo 3.3, figura 1). Debido a que la recombinación entre las dos clases cromosómicas estará reducida en los heterocigotos para la inversión, las regiones incluidas podrán evolucionar de manera independiente. No obstante, el aislamiento no es total, ya que puede existir un cierto flujo genético entre las diferentes inversiones mediante dobles entrecruzamientos o eventos de conversión génica.

Los niveles de variación nucleotídica estimados para la región OS de *D. subobscura* ($\pi_{SIL} = 0.013$) son similares a los estimados para la misma región en *D. simulans*, y por lo tanto, mucho mayores que los de *D. melanogaster*. La distribución del polimorfismo a lo largo de la región OS también se asemeja al de *D. simulans*, siendo de nuevo el gen OS-F el menos variable con la excepción de una región de más alta variabilidad en su primer intrón (capítulo 3.3, figura 3). En todas las especies donde se ha estudiado el polimorfismo se han detectado regiones altamente conservadas en el primer exón y primer

intrón del gen *OS-F*, no presentes en el gen *OS-E*, que podrían corresponder a importantes regiones reguladoras. El análisis de estas regiones conservadas (tanto en su grado de conservación -a nivel interespecífico-, como en el patrón de variabilidad nucleotídica intraespecífica) en otras especies del género *Drosophila* podría aclarar su papel en la regulación de los genes *OS-E* y *OS-F* y quizás en una posible diversificación funcional.

El análisis de las regiones codificadoras de los genes *OS-E* y *OS-F* en *D. subobscura* revela un desvío significativo de la relación esperada, bajo el modelo neutro, entre el número de sustituciones sinónimas y no sinónimas polimórficas, y fijadas con relación a *D. guanche* (*MK test*, $P = 0.03$). Al analizar por separado las zonas codificadoras de cada gen se observa que el desvío se produce únicamente en el gen *OS-E* ($P = 0.005$). Para intentar averiguar el sentido del desvío se analizaron estas sustituciones pero utilizando la secuencia del gen *OS-E* de *D. pseudoobscura* como grupo externo. En este caso, el resultado del test no fue significativo. Por lo tanto, los resultados del test *MK* nos indicarían la existencia de un exceso de sustituciones no sinónimas fijadas entre *D. guanche* y *D. subobscura* (y probablemente en el linaje de *D. guanche*), y no por un defecto en los niveles de polimorfismo en *D. subobscura*.

D. guanche es una especie endémica de Canarias y que probablemente presente un tamaño efectivo poblacional mucho menor que el de una especie de distribución paleártica como *D. subobscura*. Esta disminución del tamaño efectivo, desde el ancestro común a ambas especies, podría provocar un incremento de la tasa de fijación de mutaciones ligeramente deletéreas (como podrían ser las sustituciones no sinónimas) debido a una reducción en la eficacia de la selección purificadora (LLOPART *et al.* 1999; PÉREZ *et al.* 2003). Si esta hipótesis fuera cierta, ambos genes deberían verse afectados, y en principio con una magnitud similar, característica que no observamos. A pesar de que el test *MK* sólo es significativo en el gen *OS-E*, no podemos descartar, totalmente, que el gen *OS-F* también se haya visto afectado por la relajación de la selección; de hecho como este gen tiene una tasa evolutiva más pequeña acumulará menos sustituciones y presentará una menor potencia estadística. Además, las mutaciones sinónimas, en el caso de existir algún tipo de selección en contra de codones no preferentes (codones no utilizados

preferencialmente para un determinado amino ácido) en el ancestro de *D. subobscura* y *D. guanche*, también reflejarían esta reducción, incrementando su tasa de fijación. Debido a que el número de sustituciones sinónimas fijadas entre *D. subobscura* y *D. guanche* es muy reducido, no se puede realizar un análisis fino de la tasa de fijación de codones preferentes y no preferentes, al menos, con una potencia estadística aceptable.

Otra posibilidad es que el exceso de reemplazamientos fijados en el gen *OS-E* de *D. guanche* sea el resultado de la fijación de mutaciones selectivamente favorables. El análisis de la ratio $\omega = K_A/K_S$ a lo largo de la región codificadora del gen *OS-E* mediante *sliding window* muestra un pico de ω en el exón 3, donde alcanza una proporción de cambios no sinónimos (K_A) de hasta dos veces la de sinónimos, que podría ser el resultado de la acción de la selección positiva (ver capítulo 3.4). El error estocástico asociado al análisis de *sliding window*, no obstante, puede ser importante, por lo que se debe de comprobar si existe realmente alguna región del gen *OS-E* que pueda haber sufrido la acción de la selección direccional positiva; la metodología más adecuada para ello es mediante el análisis por máxima verosimilitud utilizando modelos apropiados de codones (ver sección siguiente).

El análisis conjunto del polimorfismo cromosómico y nucleotídico muestra que las dos ordenaciones cromosómicas estudiadas en este trabajo presentan una fuerte diferenciación genética ($S_{nn} = 0.965$, $P < 0.001$; HUDSON 2000). La supresión (o reducción) de la recombinación en los heterocigotos para estas ordenaciones es probablemente la causa de dicha diferenciación. El aislamiento genético, no obstante, no es completo: las dos ordenaciones comparten más mutaciones (50 polimorfismos compartidos) que las esperadas por un origen independiente de las mismas en cada ordenación (2.67 ± 1.6). Además, se pudieron identificar 6 pequeños tractos de conversión génica (BETRÁN *et al.* 1997). A pesar del intercambio genético, el árbol filogenético de las secuencias de la región OS es claramente compatible con un origen monofilético de las ordenaciones ya que las secuencias de cada ordenación se agrupan en un mismo grupo. Aunque una de las líneas (TB132) se posiciona exteriormente a los dos grupos, podría representar algún intercambio genético

(doble entrecruzamiento o conversión génica) de este cromosoma con el de otras ordenaciones no analizadas en el presente estudio.

Las estimas de diversidad nucleotídica silenciosa de la ordenación $O_{[3+4]}$ ($\pi_{SIL} = 0.014$) son mayores que los de la ordenación con la inversión 23 ($\pi_{SIL} = 0.009$). El patrón de distribución de la variabilidad nucleotídica, no obstante, es similar en ambas clases y presenta valores negativos de los estadísticos de TAJIMA (1989) y de FU and LI (1993). Aunque estos valores no son significativos –asumiendo que la recombinación intragénica es nula- si incorporamos en las simulaciones de coalescencia el valor de recombinación más conservativo (estimado independientemente en cada una de las ordenaciones), los resultados de los test son altamente significativos.

Probablemente, el patrón de distribución de las mutaciones en cada una de las dos ordenaciones refleja aún el efecto del proceso de expansión producido por el incremento de frecuencia de la inversión (desde su origen) hasta alcanzar su frecuencia actual (quizás la de equilibrio). Un proceso de expansión poblacional genera una distribución del número de diferencias entre pares de secuencias (*mismatch distribution*) característica (muy similar a la distribución de Poisson; SLATKIN and HUDSON 1991; ROGERS and HARPENDING 1992). No obstante, en presencia de recombinación una población de tamaño constante (estado estacionario) puede presentar una *mismatch distribution* que también se asemeje a una distribución de Poisson. Así pues, la forma de la distribución (como el resultado de los estadísticos basados en ella) no es una evidencia de un proceso de expansión reciente. Los estadísticos R_2 y F_s (FU 1997; RAMOS-ONSINS and ROZAS 2002) son mucho menos sensibles al efecto de la recombinación intragénica y permiten detectar la huella dejada por un proceso reciente de expansión. Los valores de estos estadísticos son significativos en las dos ordenaciones cromosómicas estudiadas. Por lo tanto, el patrón de distribución de la variabilidad nucleotídica intraespecífica posiblemente está reflejando el efecto del proceso de expansión poblacional que afectó a cada una de las inversiones cromosómicas a partir de su origen. Consecuentemente, la información nucleotídica de las dos ordenaciones cromosómicas aún no ha alcanzado el equilibrio.

Este desvío del estado estacionario nos puede permitir estimar la edad de la inversión cromosómica O_{23} mediante el análisis del patrón de variabilidad nucleotídica que existe dentro de ella (ROZAS *et al.* 1999). Asumiendo que la inversión O_{23} tiene un origen único, y que por tanto toda la variación presente en esta inversión se originó con posterioridad (a excepción de la incorporada por flujo genético -conversión génica o dobles entrecruzamientos-), y que las secuencias aún no han llegado al equilibrio (genealogía en estrella), se puede estimar el tiempo de coalescencia de todas las secuencias que en el fondo será una estima de la edad de la propia inversión. El análisis requiere de la estima de la tasa de mutación (estimada a partir de los datos de divergencia), y de datos de la variabilidad intraespecífica. A partir de la estima de divergencia silenciosa entre *D. subobscura* y *D. guanche* ($K_{SIL} = 0.079$) y asumiendo un tiempo de divergencia de entre 1.8 y 2.8 millones de años para estas dos especies (RAMOS-ONSINS *et al.* 1998), la tasa de sustitución nucleotídica silenciosa sería de entre 2.19×10^{-8} y 1.41×10^{-8} sustituciones por sitio y por año. Con estos valores la edad de la inversión O_{23} sería de 0.19 a 0.29 millones de años. Estos valores son algo menores que los estimados para la ordenación O_{3+4+23} a partir de la variabilidad en el gen *rp49* (0.28-0.43 millones de años) (ROZAS *et al.* 1999). En este último trabajo, sin embargo, la posición de la región de estudio (*rp49*) estaba más alejada del punto de rotura de la inversión cromosómica, y podría haber incorporado más información genética –por flujo génico- procedente de otras ordenaciones cromosómicas.

4.3. Evolución molecular de los genes *OS-E* y *OS-F* en el subgénero *Sophophora*

Se ha realizado un análisis de variabilidad en las zonas codificadoras de los genes *OS-E* y *OS-F* en 14 especies pertenecientes al subgénero *Sophophora* de *Drosophila*: *D. melanogaster*, *D. simulans*, *D. mauritiana*, *D. yakuba*, *D. teissieri*, *D. erecta* y *D. ananassae*, del grupo *melanogaster* y *D. subobscura*, *D. guanche*, *D. madeirensis*, *D. pseudoobscura*, *D. miranda*, *D. persimilis* y *D. bifasciata*, del grupo *obscura*.

La secuencia de DNA de la región codificadora de las especies fue obtenida mediante amplificación por PCR usando oligonucleótidos diseñados en las regiones más conservadas de las secuencias de las especies más distantes ya determinadas. Para completar la secuenciación se utilizaron oligonucleótidos específicos para cada especie. Los análisis se realizaron usando un alineamiento múltiple específico para cada gen -con toda la secuencia codificadora, e incluyendo péptido señal-, y un tercer alineamiento con las secuencias correspondientes a la proteína madura (sin incluir el péptido señal) de los dos genes. Para el análisis de la variabilidad nucleotídica (región codificadora) se utilizaron métodos de máxima verosimilitud usando la información filogenética y un modelo específico de codones (GOLDMAN and YANG 1994; YANG 2000; KOSAKOVSKY POND and FROST 2005).

En todas la especies estudiadas –tanto del grupo *melanogaster* como del *obscura*- se identificaron los ortólogos de los genes *OS-E* y *OS-F*, mientras que en *D. virilis* (al igual que HEKMAT-SCAFE *et al.* 2000) no se pudo identificar el ortólogo del gen *OS-E*. Esta observación es compatible con nuestra estima de la edad de la duplicación a partir de la información de la variabilidad nucleotídica (43-94 millones de años, asumiendo que la separación entre los grupos *melanogaster* y *obscura* dentro del subgénero *Sophophora* ocurrió hace unos 25-55 millones de años, RUSSO *et al.* 1995; TAMURA *et al.* 2004). En todas las especies los genes *OS-E* y *OS-F* se sitúan en *tandem*, con la misma orientación, con un tamaño muy similar y con un patrón de distribución intrón-exón equivalente. La distancia física entre los genes, a pesar de ser variable, también se mantiene razonablemente conservada. La estructura de ambos

genes, por lo tanto se ha mantenido relativamente estable durante un largo periodo de tiempo, lo que sugiere la existencia de algún importante mecanismo de preservación.

Los genes *OS-E* y *OS-F* presentan una marcada diferencia en la tasa evolutiva global, tanto sinónima como no sinónima, que puede reflejar una diversificación funcional de estos genes. Para intentar averiguar si algún tipo de diversificación funcional pudiera estar implicado en el mantenimiento de la duplicación se analizaron los niveles de constricción funcional de cada uno de los genes. Los niveles de divergencia sinónima y no sinónima en parejas de secuencias indican que el gen *OS-E* evoluciona mucho más rápidamente que el gen *OS-F* (capítulo 3.4, figura 3). De hecho, el gen *OS-E* (región codificadora) acumula un número significativamente mayor de sustituciones nucleotídicas que el gen *OS-F* (RRT, $P = 0.001$). No obstante, tanto el análisis de la divergencia nucleotídica, como el de máxima verosimilitud muestran que la constricción funcional es alta y aproximadamente la misma en los dos genes ($\omega \sim 0.06$). Estos resultados concuerdan con los obtenidos para duplicaciones antiguas (LINCH and CONERY 2003a, 2003b), y no descarta que hubiera procesos de diversificación funcional en los momentos inmediatamente posteriores a la duplicación. La distribución de la divergencia sinónima y no sinónima no es homogénea a lo largo de la zona codificadora de estos genes, destacando dos regiones altamente variables para ambos tipos de sustituciones (regiones *het1* y *het2*; capítulo 3.4, figura 9). Así, las distintas tasas evolutivas de *OS-E* y *OS-F* podrían reflejar algún tipo de diferenciación que haya contribuido a la preservación de la duplicación.

El análisis de máxima verosimilitud indica que los datos se ajustan mejor a un modelo donde los dos genes evolucionan con la misma constricción funcional en todos los linajes (con excepción de *D. guanche*). De todas maneras, la constricción funcional no es homogénea a lo largo de la región codificadora; de hecho, los dos genes se ajustan mejor a un modelo con una variación de ω a lo largo de los sitios, que a un modelo con un único valor de ω para todos ellos (capítulo 3.4, tabla 1). En particular, el modelo que mejor se ajusta a los datos asume dos clases de sitios con diferente presión selectiva,

aunque siempre con valores muy bajos de ω y sin evidencias de sitios que hayan evolucionado por selección positiva.

Sólo aquellos modelos que permiten la inferencia de la selección positiva en sitios particulares de linajes concretos (modelos de *branch-site*), mostraron un resultado significativo (capítulo 3.4, figura 7); concretamente, en el linaje de *D. guanche*, en el de la rama más interna que separó a ambos genes, y en los linajes que conducen a las especies del grupo melanogaster en cada uno de los 2 genes (capítulo 3.4, tabla 2). Los resultados de este análisis permitieron identificar 15 posiciones seleccionadas favorablemente. Interesantemente, 5 de estas posiciones implican cambios de amino ácido de tipo radical (en relación a la carga o a la polaridad) fijados entre los dos genes. Estos cambios, promovidos por la selección natural positiva, podrían haber tenido un papel relevante en la diversificación funcional de los genes *OS-E* y *OS-F*, y por ende en el mantenimiento de ambas copias en el genoma de forma estable.

Se ha determinado la posición de los residuos aminoacídicos potencialmente interesantes en la putativa estructura tridimensional de la OBP codificada por los genes *OS-E* y *OS-F*. Al no disponerse de la estructura real de estas proteínas, se posicionaron los amino ácidos identificados en el análisis bayesiano sobre la estructura tridimensional de la OBP de *Apis mellifera* (AmelASP); esta OBP es la proteína con mayor similitud de secuencia de amino ácidos con las proteínas OS y que tiene la estructura tridimensional resuelta. Tanto los 5 amino ácidos fijados entre los genes, como la mayoría del resto de amino ácidos identificados como posibles dianas de la selección positiva, se concentran en únicamente tres segmentos de la proteína, el extremo N-terminal, formado esencialmente por la primera hélice α (donde se encuentra la región *het2*), la región formada por las hélices D y E (que incluyen casi exclusivamente la región *het1*), y el extremo C-terminal de la proteína. Es interesante destacar que estudios estructurales en OBPs con la estructura resuelta (LARTIGUE *et al.* 2003; TEGONI *et al.* 2004) han demostrado que reemplazamientos de amino ácidos en estos segmentos pueden provocar cambios importantes en las propiedades cinéticas y de solubilidad de la proteína, y en las de interacción y unión con el ligando.

Es importante matizar que alguna de las metodologías utilizadas han sido criticadas por provocar falsos positivos sobre la inferencia de la selección positiva (Zhang 2004). No obstante, en el presente trabajo se han utilizado las últimas implementaciones de los modelos de codones (ver YANG *et al.* 2005), que intentan evitar precisamente este comportamiento liberal. Además, y para evitar estimas subóptimas (estimación por máxima verosimilitud de valores de los parámetros por debajo del máximo real), se realizaron múltiples ejecuciones para cada uno de los modelos. Asimismo, se utilizó un test específico de selección muy conservativo (test 2), para evitar confundir un proceso de relajación total de la selección (selección purificadora), con el de la acción de la selección positiva. Esta última hipótesis no pudo ser rechazada en el caso de la rama interna que separa ambos genes (branch *d*; capítulo 3.4, figura 7) y en el linaje de *D. guanche*, no pudiéndose –por tanto- demostrar la existencia de selección positiva en estos linajes. Estas mutaciones identificadas como selectivamente favorables podrían en realidad ser mutaciones deletéreas complementarias (modelo DDC; FORCE *et al.* 1999; LYNCH and FORCE 2000). En cambio en las ramas *me* y *mf* los resultados del test 2 son significativos, lo que indicaría que la selección positiva pudo haber actuado en estos linajes.

Asimismo, la posible diversificación funcional de estos genes podría haber ocurrido por cambios en regiones no codificadoras. En *D. melanogaster*, estos genes tienen un mismo patrón espacial de expresión génica; por lo tanto, no parece que la diversificación se dé a este nivel. Por otro lado, al no existir datos experimentales sobre los niveles de expresión relativos de estos genes, no se puede determinar si existen diferencias a este nivel de regulación. En este contexto, la diferente estructura génica de los dos genes (presencia de un exón no traducido y de un extenso intrón en el gen *OS-F* y no presentes en el *OS-E* capítulo 1, figura 1) puede tener importancia. Los análisis de polimorfismo y divergencia en *D. melanogaster*, *D. simulans* y *D. subobscura* han identificado dos regiones altamente conservadas dentro de dicho intrón, candidatas a ser importantes regiones reguladoras. En este sentido, recientemente se ha argumentado que el tamaño del primer intrón está correlacionado positivamente con los niveles de expresión, efecto que podría estar causado por la presencia de regiones reguladoras en dichos intrones (MARAIS *et al.* 2005). Por lo tanto, las diferencias en la estructura génica

podrían provocar diferencias en los niveles de expresión que podrían ser determinantes en la diversificación funcional de los genes *OS-E* y *OS-F*. Sería, pues, necesario cuantificar los niveles de expresión de estos dos genes, y estudiar la existencia de elementos reguladores en las regiones conservadas del primer intrón del gen *OS-F*; esta información puede ser crucial para determinar si existen diferencias susceptibles que puedan justificar el mantenimiento en el genoma de ambos miembros. Asimismo sería también interesante realizar estudios funcionales para investigar si los sitios que potencialmente han sido promovidos por la selección natural positiva tienen o han tenido un papel relevante en la función de estas OBPs.