

Patrones de variación nucleotídica y cartografía  
de bloques de selección ligada en el genoma de  
*Drosophila melanogaster*

---



# Patrones de variación nucleotídica y cartografía de bloques de selección ligada en el genoma de *Drosophila melanogaster*

*Patrons de variació nucleotídica i cartografia dels blocs de selecció lligada al genoma de Drosophila melanogaster*

*Nucleotide variation patterns and linked selection blocks mapping along the Drosophila melanogaster genome*

**TESIS DOCTORAL**

Doctorado en genética

Maite Garazi Barrón Aduriz

Director: Antonio Barbadilla Prados

**UAB**  
Universitat Autònoma  
de Barcelona

Universitat Autònoma de Barcelona  
Facultat de Biociències  
Departament de Genètica i de Microbiologia  
Bellaterra, 2015



Amari eta Aitari eta zuen amei, amona Maite eta amona Anttoniri



---

## ÍNDICE

1. Introducción	3
1.1. Evolución molecular	5
1.1.1. La teoría neutralista de la evolución molecular	6
1.2. Detección de la selección natural en los genomas	10
1.2.1. Las huellas de la selección en el genoma	10
1.2.2. Prueba de McDonald y Kreitman	13
1.2.3. Estimaciones de la selección adaptativa	17
1.2.4. Estimaciones de la selección negativa	19
1.3. Recombinación en el nivel molecular	20
1.4. Retos a la teoría neutralista	28
1.4.1. La selección ligada y el efecto de la recombinación	28
1.4.2. La selección ligada y la paradoja de la variación	34
1.5. <i>Drosophila melanogaster</i> como organismo modelo	38
1.5.1. Características del genoma	39
1.5.2. Filogenia	39
1.5.3. Demografía	40
1.6. Proyecto DGRP	42
1.7. Objetivos	46
1.7.1. Descripción y explicación de los patrones de variación en los genes codificadores de <i>Drosophila melanogaster</i>	46
1.7.2. Efecto de la recombinación sobre los patrones de variación nucleotídica	47
2. Materiales y métodos	51
2.1. Secuencias genómicas	51
2.2. Estimación de la diversidad y la divergencia	54
2.2.1. Aproximación centrada en los genes codificadores	55
2.2.2. Aproximación por ventanas consecutivas no solapantes	56
2.3. Elección del taxón externo	57
2.4. Estimación de la Recombinación	57
2.4.1. Estimaciones corregidas de recombinación	59
2.5. Estimación de la densidad génica	59
2.6. ANOVA de la variación nucleotídica y las variables genómicas	59
2.7. Determinación del umbral de recombinación	60
2.8. Análisis estadístico	61

2.9. Estimación del desequilibrio de ligamiento ( $r^2$ )	62
2.10. Estimación de la razón $c/\mu$	62
2.11. Simulaciones del efecto de la recombinación	63
2.12. Estimación de la selección	64
2.13. Software utilizado	64
3. Resultados	67
3.1. Descripción y caracterización de la variación en los genes codificadores de <i>Drosophila melanogaster</i>	67
3.1.1. Diferencias entre clases funcionales de sitios	69
3.1.2. Diferencias entre cromosomas	71
3.1.3. Diferencias entre regiones cromosómicas	75
3.2. ANOVA para determinar las variables explicativas de la variación sinónima observada	81
3.3. Valor umbral de recombinación	84
3.3.1. Comparación y validación de las estimas de recombinación en ventanas de 100kb	84
3.3.2. Correlación entre la recombinación y la diversidad (polimorfismo y divergencia)	87
3.3.3. Desacoplamiento entre polimorfismo y recombinación	88
3.4. Mapa de LSB	92
3.4.1. Patrón de LSB en autosomas y el cromosoma X	92
3.4.2. Fiabilidad del mapa de LSB	95
3.4.3. Densidad génica y el mapa de LSB	97
3.5. Características de los LSB	98
3.5.1. Razón $c/\mu$	98
3.5.2. Patrón de desequilibrio de ligamiento ( $r^2$ )	103
3.6. Correlación entre polimorfismo y divergencia en los LSB	106
3.7. Eficiencia de selección en los LSB y NLSB	108
3.7.1. Descripción general de las estimas de selección	108
3.7.2. Diferencias entre LSB y NLSB	109
3.7.3. Comparación entre LSB y NLSB situados fuera de las regiones centroméricas	110
3.8. Mapa de LSB para diversos tamaños de ventana	112
3.8.1. El mapa de LSB y NLSB	113
3.8.2. Evaluación del mapa de LSB en ventanas de 10kb	114
4. Discusión	121
4.1. Genómica de poblaciones en <i>Drosophila melanogaster</i>	121
4.2. La recombinación es el principal factor modulador de la variación nucleotídica a lo largo de los cromosomas	122



---

4.3. Bloques de selección ligada: definición, medida e implicaciones	124
4.4. Fiabilidad de los LSB estimados	126
4.5. Efecto del tamaño de ventana sobre los LSB	127
4.6. Relación entre $\rho$ y $c$	132
4.7. Eficiencia de selección	133
4.8. Singularidad del cromosoma X	136
5. Conclusiones	141
6. Bibliografía	145
7. Anexos	166
7.1. Tablas Suplementarias	166
7.2. Anexo A: Información básica de la secuenciación de las 158 líneas DGRP	223
7.3. Anexo B: Programas Perl	226
7.3.1. Fragmentar los alineamientos cromosómicos en alineamientos génicos	226
7.3.2. Eliminar sitios ambiguos	227
7.3.3. Estimaciones de polimorfismo utilizando PDA	228
7.3.4. Estimaciones de divergencia utilizando PDA	229
7.4. Anexo C: Paper DGRP	231
8. Índice de figuras	238
9. Índice de tablas	240
10. Índice de cuadros	242
11. Índice de tablas suplementarias	243



---

## PRÓLOGO

Gracias a la secuenciación de última generación (NGS) que permite disponer de múltiples genomas de individuos de una misma población podemos hoy hablar de una nueva disciplina: la genómica de poblaciones. El análisis de la variación a escala genómica añade una nueva dimensión en la interpretación de la variación genética (Jorde *et al.* 2001). Uno de los proyectos pioneros y más ambiciosos de genómica de poblaciones es la iniciativa internacional *Drosophila Genetic Reference Panel* (DGRP), en el que se han secuenciado 205 genomas de la especie *Drosophila melanogaster*. Este recurso único se ha puesto a disposición de la comunidad drosofilista para el estudio de la variación genotípica y fenotípica en este organismo modelo de la genética desde sus orígenes. Los resultados obtenidos de los recientes estudios de variación genómica han desplazado el debate neutralista-seleccionista (Lewontin 1974) hacia una nueva perspectiva: la recombinación parece ser la fuerza evolutiva clave a la hora de determinar la importancia relativa de la deriva genética *versus* la selección natural en diferentes regiones del genoma.

Esta tesis es un estudio de genómica de poblaciones en la que se describen los patrones de variación nucleotídica de una población norteamericana de *Drosophila melanogaster* usando las secuencias genómicas de DGRP. Se evalúa la importancia relativa de los posibles factores genómicos moduladores de la variación nucleotídica a lo largo del genoma. La recombinación es con mucho la principal variable explicativa, por lo que se ha explorado en profundidad cómo afecta la tasa de recombinación a los patrones de variación. La correlación universalmente hallada entre el polimorfismo y la recombinación desaparece a partir de un valor umbral de recombinación en el nivel del cromosoma. En esta tesis se demuestra y estima la existencia de este valor umbral de recombinación para cada brazo cromosómico y se introducen dos nuevos conceptos en la genómica de poblaciones: bloque de selección ligada (LSB) y bloque sin selección ligada (NLSB). El genoma es un mosaico constituido por ambos tipos de bloques en los que la dinámica de la evolución molecular es opuesta y en consecuencia deben ser reconocidos y cartografiados. En los NLSB la interpretación clásica de los datos basados en la neutralidad como modelo nulo podría adoptarse perfectamente (Cavalli-Sforza 1966; Lewontin y Krakauer 1973), mientras que en los LSB se debe incorporar la selección ligada al cuerpo teórico y los modelos al uso de la genética de poblaciones. En esta tesis se traza el mapa de LSB en la especie *Drosophila melanogaster* utilizando distintas estimas de recombinación y se intenta determinar la unidad mínima de bloque de selección ligada.



---

## PREFACE

Nowadays, sequencing of multiple individual genomes of the same population by next generation sequencing technologies (NGS) allows us to talk about a new discipline: *Population Genomics*. Genome-wide diversity analyses add an extra layer to the interpretation of genetic variation (Jorde *et al.* 2001). A pioneer and ambitious project in population genomics is the international initiative *Drosophila Genetic Reference Panel* (DGRP), where 205 individual genomes of the species *Drosophila melanogaster* have been sequenced. This unique resource is being shared by the *Drosophila* research community to study genotypic and phenotypic diversity in this model organism. The results obtained by the new genome-wide variation studies have shifted the neutralist-selectionist debate (Lewontin 1974) towards a new perspective: the recombination is the key evolutionary force determining the relative importance of genetic drift *versus* natural selection in different regions of the genome.

In this dissertation, we describe the nucleotide diversity patterns of a North American natural population of *Drosophila melanogaster* population using DGRP genome sequences. We assess the relative importance of different population genetics forces modulating genome-wide nucleotide diversity. Recombination is the main explanatory variable and hence, we have deeply explored how recombination affects diversity patterns. The correlation almost universally found between polymorphism and recombination vanishes when recombination exceeds a given threshold value at the chromosome level. We show the actual existence of a recombination threshold at any given chromosome and their values for each chromosome arm have been estimated. Two new concepts in population genomics are introduced: linked selection block (LSB) and no linked selection block (NLSB). The genome is a mosaic made by these two distinctive blocks on which molecular evolutionary dynamics are opposite and, in consequence, they have to be recognized and mapped. In NLSB the classical interpretation of data based in neutrality as a null model could be perfectly applied (Cavalli-Sforza 1966; Lewontin y Krakauer 1973), while in LSB linked selection should be included in the theoretical framework of population genomics. We mapped the LSB map in *Drosophila melanogaster* using several recombination variables and we try to determine the minimum unit of linked selection block.



---

# Parte 1

## Introducción





---

## 1. Introducción

*“Nada en evolución tiene sentido excepto a la luz de la genética de poblaciones”\**

---

La obra *El origen de las especies* de Charles Darwin (*On the Origin of Species*, 1859), es un largo argumento (Mayr 1991) a favor de la visión de la evolución biológica como el proceso histórico que ha dado lugar a la diversidad de la vida en nuestro planeta. Darwin además propone el mecanismo de la selección natural para explicar la adaptación de las especies a su entorno. Como en cada generación nacen más individuos de los que pueden sobrevivir, cualquier variación que contribuya diferencialmente a la supervivencia de un individuo respecto a otro tenderá a propagarse según “el principio fuerte de la herencia”. La aceptación de estas dos ideas revolucionarias implica una nueva forma de comprender la variación natural: *el pensamiento poblacional* (Mayr 1963; Mayr 1976). Bajo esta visión, la variación genética y fenotípica existente entre individuos dentro de las poblaciones se amplificará en el tiempo y el espacio hasta generar nuevas poblaciones, nuevas especies y por extensión toda la diversidad de la tierra, el producto acumulativo de la evolución biológica (Lewontin 1974).

La genética de poblaciones proporciona el marco teórico para explicar la evolución biológica desde el paradigma de la variación. La unidad de evolución es la población mendeliana, que se define mediante la distribución de los alelos y los genotipos de los individuos que la forman. La evolución es en esencia el resultado de los cambios acumulativos de estas distribuciones en el tiempo. La genética de poblaciones ha desarrollado un extenso cuerpo teórico que describe la dinámica de las distribuciones de los alelos y los genotipos en poblaciones mendelianas, empezando por un estado en el que no actúa ninguna fuerza (el principio de Hardy-Weinberg) hasta considerar el impacto de distintas fuerzas en las distribuciones genéticas de las poblaciones (Wright 1931).

Hay al menos cinco fuerzas evolutivas que actúan sobre la variación genética de las poblaciones: la mutación, la deriva genética, el flujo genético, la recombinación y la selección natural. La intensidad relativa, direccionalidad y variación a lo largo del tiempo de cada una de estas fuerzas definen la forma en la que la evolución sucede en un contexto determinado (Lynch 2007). Las mutaciones son sucesos aleatorios respecto a su efecto sobre la eficacia biológica (*fitness*), es decir, que ocurren independientemente de si benefician o perjudican a los individuos en su adaptación al medio. A pesar de que muchas mutaciones se pierden en las primeras generaciones, algunas pueden incrementar su frecuencia en la población bien por estar asociadas a una mayor eficacia biológica o bien por azar, por deriva genética. Estas variantes

---

\* Frase de Michael Lynch (2007), reformulación de la frase clásica de T. Dobzhnasky: *Nada en biología tiene sentido excepto a la luz de la evolución.*

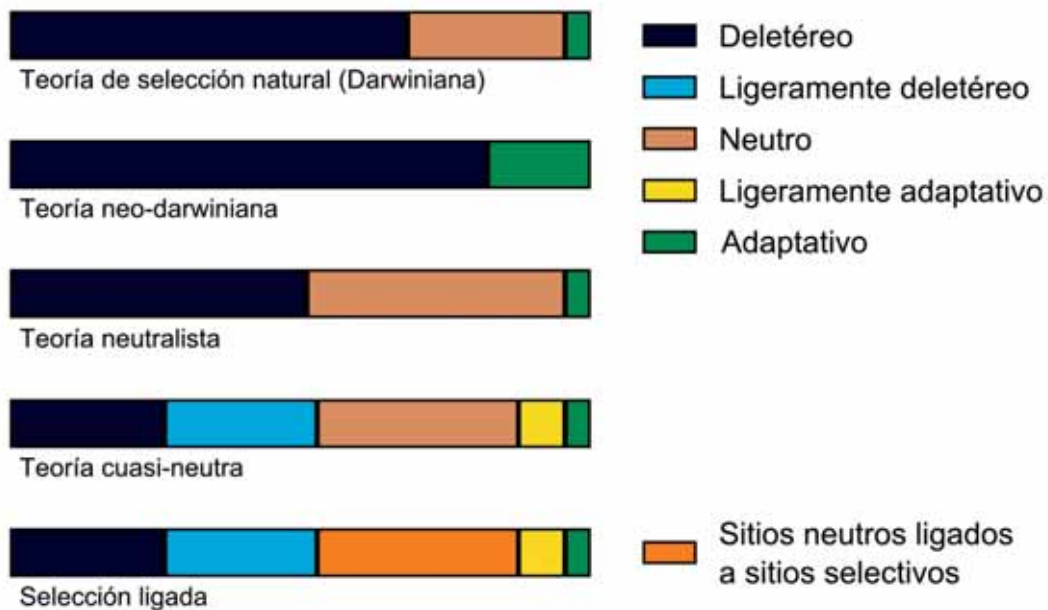
---

alélicas constituyen la variabilidad intra-poblacional que se conoce como polimorfismo, es decir, cualquier variante en la secuencia de DNA que esté presente en la población en frecuencia apreciable. Cuando un alelo polimórfico incrementa en frecuencia hasta reemplazar el resto de los alelos de la población se habla de fijación o sustitución alélica. Cuando dos poblaciones se separan o aíslan reproductivamente, evolucionan bajo procesos independientes que llevan a la fijación de diferentes alelos, cuya acumulación paulatina en el tiempo contribuye a la diferenciación de dichas poblaciones en diferentes especies. Este proceso se conoce como divergencia. El polimorfismo y la divergencia suministran información complementaria de la historia genética de las poblaciones. Mientras que el polimorfismo proporciona una información detallada de los sucesos poblacionales recientes, la divergencia abarca una ventana temporal mayor, desde la separación de la especie externa que se usa de referencia (*outgroup*) hasta el presente. Por tanto, el análisis combinado del polimorfismo intra-específico y la divergencia inter-específica es una de las aproximaciones que mayor información suministran para entender el impacto de las distintas fuerzas evolutivas sobre los patrones observados de variación genética.

La genética de poblaciones es una disciplina que tiene como principal objetivo la descripción e interpretación de dos aspectos principales de la evolución biológica: i) el mantenimiento de la variación genética dentro de las especies y ii) la conversión de la variación dentro de las especies en variación entre especies (Dobzhansky 1937). A principios de los años 30 del pasado siglo, Fisher (1930), Haldane (1932) y Wright (1931) pusieron los fundamentos matemáticos de la genética poblaciones. Desarrollaron, de forma cuantitativa, las consecuencias del muestreo aleatorio de gametos en cada generación y la selección en las poblaciones con herencia mendeliana. Posteriormente, la integración de la teoría de la genética de poblaciones junto con otros campos del estudio de la evolución como la biología experimental, la paleontología, la sistemática, la zoología y botánica, dio lugar a la teoría de síntesis la evolutiva moderna (Dobzhansky 1937, Mayr 1942, Simpson 1944, Stebbins 1950). La mayor diferencia entre esta nueva teoría, también llamada neo-Darwinista, con la teoría de la selección natural propuesta por Darwin, es la adición de las leyes de la herencia mendeliana en el marco de la teoría de la genética de poblaciones. Bajo el neo-Darwinismo, la evolución se contempla como dos procesos secuenciales: i) la generación aleatoria de la variación por mutación y/o recombinación y ii) la selección direccional de estas variantes alélicas. Es decir, que la evolución de las poblaciones naturales ocurre por la generación aleatoria de mutaciones seguidas por fuerzas selectivas que actúan sobre ellas relegando a otras fuerzas evolutivas, principalmente a la deriva genética, a una posición secundaria (Figura 1.1).

Bajo esta visión panselccionista se propusieron dos hipótesis opuestas para predecir la variación genética que había en las poblaciones naturales. El modelo clásico argumentaba que la

variación genética es poco común y está principalmente compuesta por alelos deletéreos mantenidos en las poblaciones mediante un equilibrio de entrada por mutación y eliminación por selección negativa (Muller y Kaplan, 1966). Consecuentemente, este modelo predice que la mayoría de los *loci* serán homocigóticos. En contraposición, el modelo equilibrado (*balanced*) propone que la selección a favor de los heterocigotos es ubicua, y predice que la variación genética en las poblaciones naturales es abundante (Dobzhansky, 1955).



**Figura 1.1.** Importancia relativa de los distintos tipos de selección natural propuestas en las principales teorías evolutivas desde de Darwin a la actualidad

Darwin postuló la existencia de variantes deletéreas, ventajosas y neutras (Darwin 1859, Bernardi 2007). La variación neutra fue reintroducida con un protagonismo principal por Kimura (1968, 1983), a través de su teoría neutralista de la evolución molecular. La teoría cuasi-neutra, propuesta por Ohta (1992, 2002), amplía el espectro de variantes existentes en las poblaciones introduciendo las variantes ligeramente seleccionadas, tanto positiva como negativamente. En los últimos años, se ha puesto de manifiesto la importancia de la selección ligada y se asume que implícitamente afecta a todo el genoma (Hahn 2008). Figura modificada de Bernardi (2007), Ohta (2002) y Bromham y Penny (2003).

## 1.1. Evolución molecular

Tras la clarificación de la estructura en doble hélice del DNA (Watson y Crick 1953) y la decodificación del código genético (Crick *et al.* 1961), una rápida sucesión de técnicas innovadoras permitieron estimar la diversidad genética por primera vez. Así en los años 60, la genética de poblaciones entró en la era molecular de la mano de los primeros estudios electroforéticos en *Drosophila* (Johnson *et al.* 1966; Lewontin y Hubby 1966) y humanos (Harris 1966), en los que se describía las diferencias en movilidad electroforética de ciertas proteínas enzimáticas como resultado de las diferencias alélicas en sus secuencias, es decir, la variación alozímica. La variación genética se empezó a medir de dos maneras: i) la proporción

---

promedio de los *loci* que son heterocigotos en un individuo (heterocigosidad o diversidad génica) y ii) la proporción promedio de *loci* que son polimórficos en la población (el polimorfismo). Es decir, que por primera vez en genética de poblaciones se pudieron obtener medidas directas de variación genética de un gran número de especies, permitiendo que esta ciencia pasara de ser casi exclusivamente teórica a disponer de datos empíricos con los que poner a prueba los modelos teóricos. Este tipo de estudios revelaron una gran variabilidad genética en la mayoría de las poblaciones de las especies analizadas (Nevo *et al.* 1984), muy superior a la predicha teóricamente hasta ese momento.

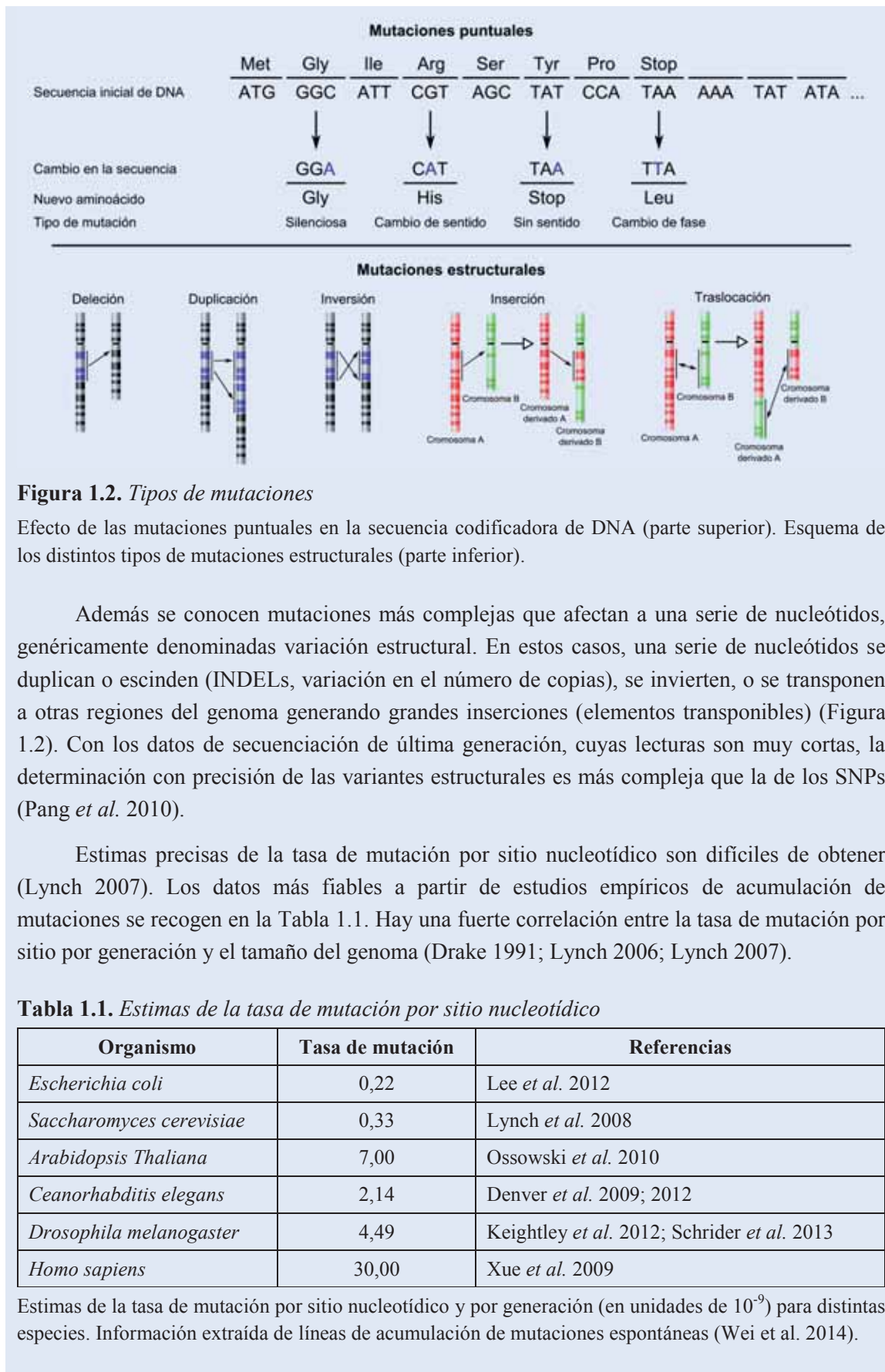
Este hecho permitió probar las hipótesis surgidas bajo la teoría neo-darwinista: la clásica y la equilibradora. Aunque en un primer momento el modelo equilibrado parecía vindicado respecto del clásico, el exceso de variación encontrada propició la propuesta de una teoría alternativa: la teoría neutralista de la evolución molecular del genetista japonés Motoo Kimura (1968), que postula que la mayoría de las sustituciones nucleotídicas son el resultado de la fijación aleatoria de mutaciones selectivamente neutras en lugar de ser mantenidas como un polimorfismo equilibrado como proclamaba el modelo equilibrado (Figura 1.1).

### 1.1.1. La teoría neutralista de la evolución molecular

Esta teoría postula que los niveles de variación genética están determinados únicamente por la entrada de alelos nuevos a través de la mutación (Cuadro 1) y la pérdida o fijación de alelos por deriva genética. Este proceso se conoce también como el equilibrio mutación-deriva. Debido a la aleatoriedad de la mutación, una fracción significativa de los nuevos alelos alterará la función de los genes y serán deletéreos, sufriendo los efectos de la selección negativa, y el resto de las mutaciones serán neutras. La teoría neutralista no niega la selección adaptativa, pero las mutaciones favorables, que ocurren en el transcurso de la evolución, son tan raras y se fijan o pierden tan rápidamente que no suelen contribuir al polimorfismo de las poblaciones.

#### Cuadro 1. Mutación

Una mutación es un cambio estable y heredable en el material genético. Las mutaciones son la fuente primaria de la variación genómica y constituyen la materia prima de la evolución. A las mutaciones que modifican un único nucleótido en la secuencia DNA se les denomina mutaciones puntuales. Si estas mutaciones superan el 1% de frecuencia en la población se suelen considerar SNPs (*Single Nucleotide Polymorphism*). Los SNPs pueden ocurrir a lo largo de todo el genoma pero las que se encuentran en las regiones codificadoras pueden diferenciarse entre SNPs sinónimos, que no modifican la cadena aminoacídica, o SNPs no sinónimos que generan un cambio de aminoácido en la secuencia proteica (Figura 1.2).



---

Como toda mutación aparece inicialmente en un individuo, la naturaleza estocástica de la transmisión mendeliana hace que la gran mayoría de las mutaciones sean eliminadas, incluso si son ligeramente ventajosas, en las primeras generaciones. Sólo una pequeña parte se propagará por la población hasta su fijación, y esto requerirá un tiempo relativamente grande (un promedio de  $4N_e$  generaciones en el caso de las variantes neutras, Kimura y Ohta 1969). Esta visión se recoge en el modelo clásico de Wright-Fisher, en el que la trayectoria de los alelos se considera como un proceso hacia adelante en el tiempo, desde el origen por mutación de un alelo hasta su pérdida o fijación en la población, y ésta ha sido la perspectiva dominante de la genética de poblaciones durante 50 años (Hartl y Clark 2007). Sin embargo, es igual de válido considerar este proceso yendo hacia atrás en el tiempo. Al reconstruir el pasado desde el presente, se pueden obviar las mutaciones que se han perdido en el camino, que son la inmensa mayoría, por lo que esta aproximación es mucho más expeditiva y su modelización más eficiente (Cuadro 2).

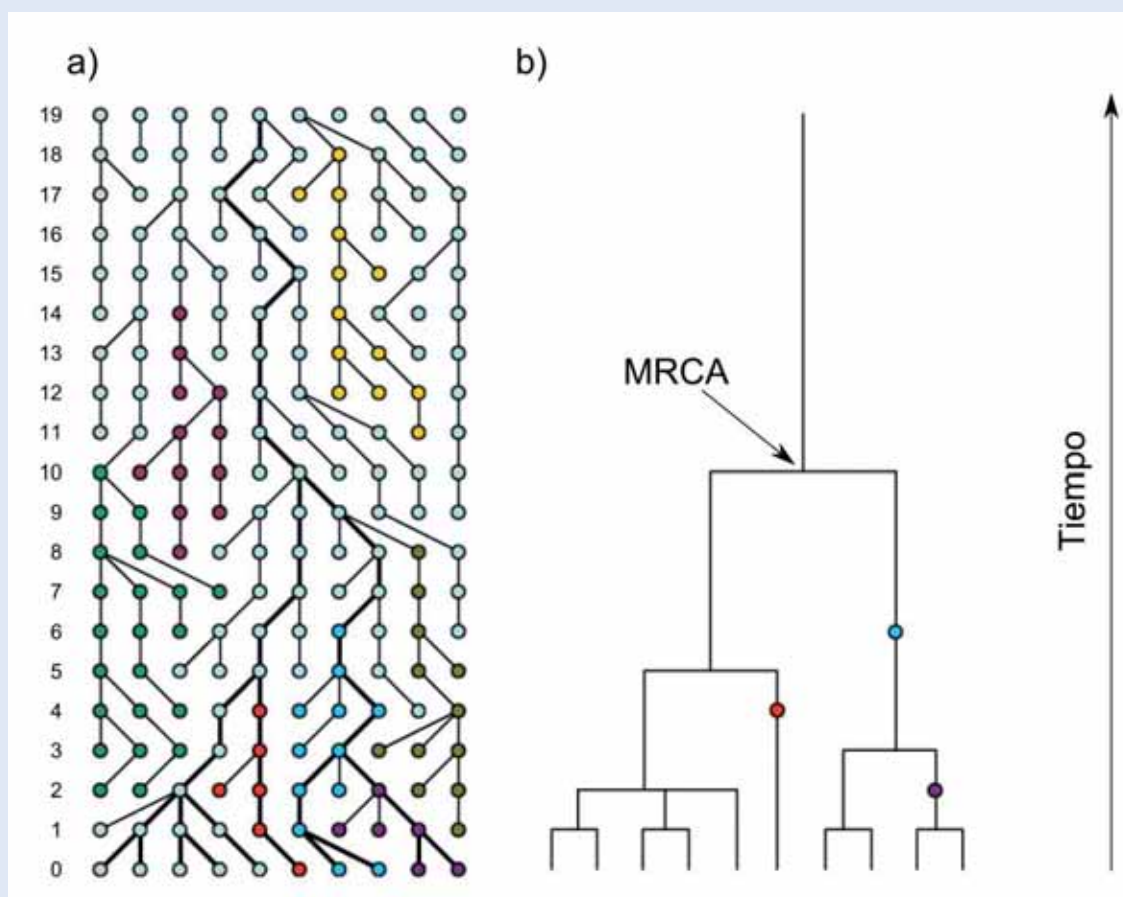
Para probar este modelo teórico se requieren estimas precisas de la diversidad nucleotídica y la divergencia. Bajo el equilibrio de mutación-deriva, la heterocigosidad esperada entre dos secuencias dadas de una misma especie ( $\theta$ ) viene dada por la expresión de  $4N_e\mu$ , donde  $N_e$  es el tamaño efectivo de la población y  $\mu$  es la tasa de mutación neutra (Tajima 1983). La divergencia nucleotídica es el número promedio de diferencias nucleotídicas medidas entre dos secuencias de especies diferentes ( $k$ ). Si la divergencia es un producto de la fijación de los alelos neutros, ésta se equipara con la expresión  $2T\mu$ , donde  $T$  es el tiempo transcurrido desde la separación del ancestro común de las dos secuencias (Kimura, 1968).

Teniendo esto en cuenta, la teoría neutralista hace dos importantes predicciones: (i) Las poblaciones con mayor censo efectivo tendrán un mayor nivel de variación genética (heterocigosidad). Por lo tanto, debería observarse una correlación lineal entre  $N_e$  de las poblaciones y sus niveles de  $\theta$ . Sin embargo,  $k$  es independiente de  $N_e$  y esto permite estimar el tiempo relativo de divergencia entre especies (el llamado reloj molecular). (ii) Debe existir una correlación positiva entre la variación inter e intra específica debido a que tanto el polimorfismo como la divergencia dependen linealmente de la magnitud de la tasa de mutación neutra ( $\mu$ ).

A pesar del largo debate de más de 40 años sobre la validez de la teoría neutralista (Kimura 1983, Gillespie 1991), ésta se ha convertido en el marco teórico por excelencia de la genética de poblaciones y la evolución molecular. Independientemente de cómo y cuánto actúan las otras fuerzas evolutivas, la hipótesis neutralista proporciona una base teórica para la comprensión de la variación del DNA con hipótesis claras y comprobables además de un conjunto de herramientas estadísticas para distinguir la selección natural de la deriva genética (Kreitman 2000, Nielsen 2001, Hahn 2007).

## Cuadro 2. La teoría de la coalescencia

La teoría de la coalescencia proporciona un modelo estocástico simple para la historia de una muestra genética siguiendo el modelo de Wright-Fisher. Originalmente ideado por Kingman (1982a, 1982b), se ha convertido en el modelo teórico por excelencia para el análisis de la variación de secuencias de DNA en la genética de poblaciones. El término coalescencia se refiere al proceso en el que mirando hacia atrás en el tiempo las genealogías de los alelos se unen en un ancestro común. Es decir, mediante este modelo se intenta relacionar todos los alelos de un gen compartidos por todos los individuos de una población a una única copia ancestral conocida como el ancestro común más cercano (MRCA, *Most Common Recent Ancestor*). Las relaciones de herencia entre alelos se representan típicamente mediante un árbol de coalescencia, o “genealogía génica”, muy similar a un árbol filogenético (Figura 1.3). Comprender las propiedades estadísticas de estas genealogías bajo diferentes supuestos constituye el núcleo de la teoría de la coalescencia.



**Figura 1.3.** Esquema de la simulación de una población Wright-Fisher y su correspondiente árbol de coalescencia

(a) Una muestra de una población bajo el modelo de deriva genética de Wright-Fisher con mutación. La población original se representa en la parte superior (generación 19 hacia atrás). La generación 0 es la población actual (el presente). Cada generación surge del muestreo de la generación anterior haciendo que cada alelo presente en la población puede replicarse y pasar a la siguiente generación o no dejar descendientes y desaparece de la población. (b) El árbol coalescente correspondiente. En ambas figuras cada nueva mutación se representa con un círculo de color distinto. Figura adaptada de Auton (2007).

---

Siguiendo la Figura 1.3, podemos ver que yendo de la generación 0 (presente) a la generación anterior (1) hay cuatro sucesos de coalescencia, es decir, que 8 alelos actuales “coalescen” de dos en dos en un cuatro únicos alelo ancestral. A medida que vamos hacia atrás en el tiempo el número de alelos ancestrales se mantiene constante o disminuye y cada reducción en el número de alelos ancestrales es un suceso de coalescencia. Debido a que la mutación puede ocurrir durante este proceso, los alelos observados en el presente no serán idénticos en secuencia a los alelos originales, pero serán descendientes de un alelo ancestral común único. Bajo el modelo de coalescencia, el tiempo normalmente se mide en unidades de  $2N_e$  generaciones. Una de las razones por las que el razonamiento de la coalescencia es tan poderoso es que el coste computacional 'hacia atrás' en el tiempo es mucho menor que si se analiza desde una perspectiva 'hacia delante' en el tiempo, en el que muchos alelos simulados son un cálculo desperdiciado ya que no dejan descendientes en el presente. En el caso de la Figura 1.3, 164 de los 200 alelos simulados no aportarían información. Mientras que yendo hacia atrás en el tiempo sólo tendríamos que simular 36 alelos. Si imaginamos una muestra de cientos de alelos, el seguimiento de la mayoría de los linajes supondrá un coste innecesario.

## 1.2. Detección de la selección natural en los genomas

La teoría neutralista de la evolución molecular proporciona una hipótesis nula frente a la que se puede contrastar el efecto de la selección natural a partir del muestreo de datos genéticos que permitan estimar el polimorfismo, la divergencia o el grado de desequilibrio de ligamiento. La identificación de genes o regiones reguladoras que han evolucionado bajo selección adaptativa permite determinar los *loci* clave en la evolución de una especie, es decir, los *loci* con una importancia funcional relevante y distintiva. Además, la detección de regiones no codificadoras conservadas, sometidas a selección purificadora, permite a su vez descubrir y anotar regiones funcionales nuevas en los genomas. Esta capacidad para detectar las regiones importantes funcional y evolutivamente hace que la búsqueda de las huellas generadas por la selección positiva y/o la selección purificadora sea uno de los temas de investigación más prolíficos en la genética de poblaciones actual.

Por otra parte, los análisis de variación en genética de poblaciones ya no se limitan al estudio de un único gen o regiones. Desde comienzos de este siglo, con la entrada de la era de la genómica, la genética de poblaciones puede por fin analizar la variación de genomas completos de diversas especies. Este tipo de análisis de la variación a escala genómica añade una nueva dimensión en la interpretación de la diversidad genética y en la descripción detallada de la selección (Jorde *et al.* 2001).

### 1.2.1. Las huellas de la selección en el genoma

Cuando los alelos bajo selección positiva incrementan en frecuencia y se fijan o se mantienen en forma de polimorfismo estable al alcanzar una frecuencia de equilibrio en la población, dejan huellas características, patrones distintivos de variación nucleotídica en la



---

región genómica bajo selección. Uno de los retos más importantes en genética de poblaciones es poder distinguir si estas huellas observadas en el genoma son debidas a selección o a procesos demográficos tales como cuellos de botella, expansiones o subdivisiones, que pueden dejar patrones similares en la variación nucleotídica. En general, un criterio clásico para diferenciar entre ambos procesos es suponer que los procesos selectivos ocurren de forma local mientras que los procesos demográficos afectan a todo el genoma (Lewontin y Krakauer 1973).

Se han propuesto al menos cinco huellas características que deja la selección adaptativa en la variación nucleotídica (Tabla 1.2; Sabeti *et al.* 2006):

- i) *Una proporción elevada de alelos no sinónimos.* Las variantes genéticas que alteran las secuencias proteicas, las mutaciones no sinónimas, suelen ser deletéreas y no alcanzan frecuencias elevadas en la población. La selección positiva, sin embargo, puede incrementar la frecuencia de estos alelos hasta su fijación (Li *et al.* 1985; Hughes y Nei 1988). Este aumento puede detectarse comparando en secuencias homólogas de DNA entre especies la tasa de cambio no sinónimo respecto a la tasa de cambios sinónimos, u otras clases de mutaciones supuestamente neutras (la razón  $k_a/k_s$ ). Este incremento de sustituciones no sinónimas puede también detectarse comparando la razón  $k_a/k_s$  con la razón de diversidad intra-específica (polimorfismo) sinónima y no sinónima ( $\pi_a/\pi_s$ ; prueba de McDonald y Kreitman; 1991; véase sección iii); Tabla 1.2). La limitación de estos métodos es que sólo permiten detectar la selección que ocurre recurrentemente, pues tienen que ocurrir múltiples cambios selectivos para que estas diferencias dejen una señal (en forma de exceso de sustituciones no sinónimas) sobre un fondo de cambios neutros. Que se requiera la acumulación de cambios tiene la ventaja que las señales perduran en el tiempo, lo que permite detectar sucesos selectivos que han ocurrido durante un largo periodo de tiempo evolutivo.
- ii) *La eliminación o reducción de la diversidad genética en las regiones adyacentes.* Cuando un alelo adaptativo incrementa de frecuencia en la población, las variantes adyacentes ligadas pueden también llegar a fijarse. Este efecto autoestopista o de arrastre por selección positiva (*hitchhiking effect*; Smith y Haigh 1974; Kaplan *et al.* 1989; Figura 1.11), puede generar una serie de huellas características entre las que destaca la reducción de la diversidad genética en las regiones colindantes. El arrastre puede ser completo o parcial, eliminando o reduciendo la variación, respectivamente (Hermisson y Pennings 2005; Pennings y Hermisson 2006). Se han desarrollado diversos métodos para detectar las posibles huellas de los patrones de variación nucleotídica generados por los arrastres (Tabla 1.2).

**Tabla 1.2.** Métodos empleados para la detección de las huellas que deja la selección natural en las secuencias de DNA

<b>i) Proporción elevada de alelos no sinónimos</b>			
<b>Prueba</b>	<b>Técnica</b>	<b>Señal selectiva</b>	<b>Referencia</b>
$k_a/k_s$	Razón de las sustituciones sinónimas respecto a las no sinónimas	Razón elevada de sustituciones aminoacídicas respecto a las sustituciones sinónimas	Li y Wu 1985, Li 1993
<i>McDonald y Kreitman (MKT)</i>	Comparación de la tasa de variación sinónima y no sinónima entre y dentro de especies	Elevada tasa de sustituciones aminoacídicas respecto al polimorfismo	McDonald y Kreitman 1991
$k_a/k_s$ implementado en el software PAML	Estimas de máxima verosimilitud de $k_a/k_s$ permitiendo diferentes grados de selección por sitio	$k_a/k_s > 1$ en la porción de un gen	Nielsen y Yang 1998; Yang 1997
<b>ii, iii) Diversidad y espectro de frecuencias</b>			
<b>Prueba</b>	<b>Técnica</b>	<b>Señal selectiva</b>	<b>Referencia</b>
<i>HKA</i>	Razón del polimorfismo respecto a la divergencia comparando dos o más <i>loci</i>	Razón baja, relativa a la tasa de polimorfismo	Hudson <i>et al.</i> 1987
<i>Tajima's D</i>	Comparación de $\pi$ respecto al número de sitios segregantes corregido para la longitud de la secuencia	Valores negativos indican exceso de alelos raros	Tajima 1989
<i>Fay and Wu's H</i>	Comparación de alelos derivados a alta frecuencia respecto a alelos a frecuencias intermedias	Exceso de alelos derivados a alta frecuencia	Fay y Wu 2000
<b>iv) Incremento en el desequilibrio de ligamiento: diversidad y longitud haplotípica</b>			
<b>Prueba</b>	<b>Técnica</b>	<b>Señal selectiva</b>	<b>Referencia</b>
<i>Hudson Haplotype Test</i>	Detección de un subconjunto de haplotipos similares	Agrupamiento de haplotipos similares	Hudson <i>et al.</i> 1994
<i>LRH</i>	Longitud relativa del tamaño de los haplotipos	Alelos a alta frecuencia asociados con las regiones de mayor LD	Sabeti <i>et al.</i> 2002
<i>iHS</i>	Longitud relativa del tamaño de los haplotipos	Alelos a alta frecuencia asociados con las regiones de mayor LD	Voight <i>et al.</i> 2006
<b>v) Diferencias entre poblaciones</b>			
<b>Prueba</b>	<b>Técnica</b>	<b>Señal selectiva</b>	<b>Referencia</b>
$F_{st}$	Coefficiente de consanguinidad para medir diferenciación entre poblaciones	Valor elevado de $F_{st}$	Wright 1950, 1951; Lewontin y Krakauer 1973

Pruebas de selección más comunes clasificadas en función de las 5 huellas que deja la selección y que se han descrito en el texto principal (sección 1.2.1).

- 
- iii) *El incremento en la frecuencia de alelos derivados*, ligados al alelo adaptativo que es seleccionado positivamente. Los alelos derivados (en contraposición a los ancestrales) son más recientes en el tiempo y la teoría neutralista predice unas frecuencias más bajas que los ancestrales en la población (Watterson y Guesse 1977).
  - iv) *El incremento del desequilibrio de ligamiento* en la región. Si el arrastre por selección adaptativa es rápido relativo a la tasa de recombinación, se genera una región en desequilibrio de ligamiento. Así, se observarán haplotipos de mayor longitud del tamaño medio observado en el resto del genoma. Este tipo de señales generadas por los arrastres por selección positiva no duran mucho tiempo (Kim y Stephan 2002; Nielsen *et al.* 2005), debido a que la recombinación suele destruirlos rápidamente. Luego, estas aproximaciones únicamente pueden detectar sucesos adaptativos recientes e intensos. Se han desarrollado métodos específicos para detectar este tipo de huellas selectivas (Tabla 1.2).
  - v) *Diferenciación de las frecuencias alélicas entre poblaciones*. Si distintas poblaciones padecen presiones selectivas diferentes, la adaptación puede actuar de forma local cambiando la frecuencia de un alelo únicamente en una de las dos poblaciones. Cambios drásticos en las frecuencias génicas de dos poblaciones pueden indicar un *locus* sometido a selección positiva. Los métodos más utilizados que siguen este criterio se describen en la Tabla 1.2.

### 1.2.2. Prueba de McDonald y Kreitman

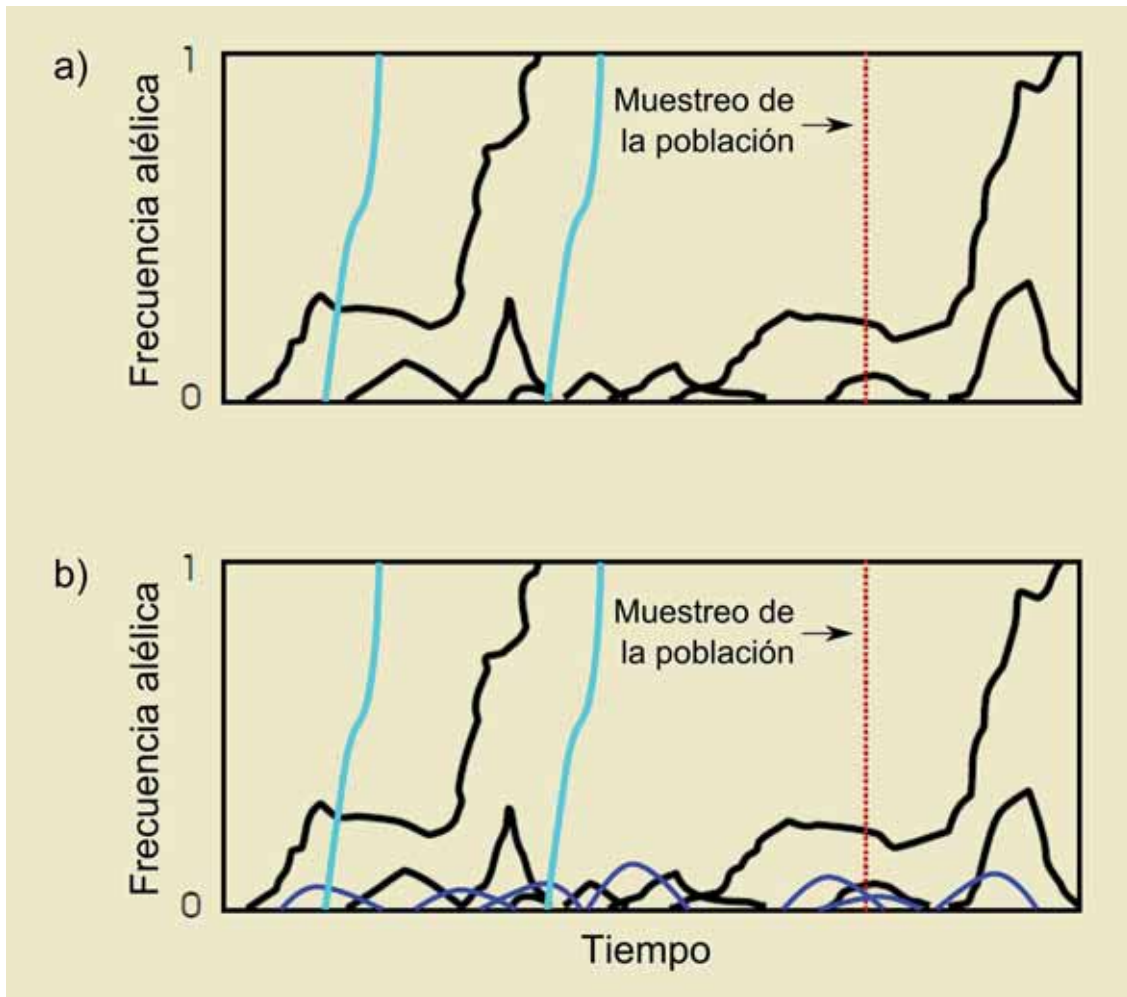
La selección natural, como se ha explicado en la sección anterior (Tabla 1.2), puede detectarse utilizando datos de divergencia, polimorfismo o la combinación de ambos. Las huellas que se pueden observar si sólo se dispone de datos de polimorfismo son débiles a menos que la selección sea reciente y relativamente intensa (Hartl y Clark 2007). Las huellas que se observan a nivel de divergencia permiten detectar diferencias acumuladas desde la separación de las especies estudiadas y, por tanto, abarcan una escala de tiempo mucho mayor, lo cual permite detectar patrones de selección en el tiempo (Hartl y Clark 2007). El uso combinado de ambos tipos de datos, polimorfismo y divergencia, no sólo permite sumar la información que suministran ambas fuentes de información, sino que proporciona evidencias adicionales en la detección de la selección adaptativa y por ello es uno de los métodos de mayor potencia para la detección de la selección.

Entre los distintos métodos existentes de inferencia de selección (Tabla 1.2) destaca la prueba o el test de McDonald y Kreitman (MKT, 1991), que combina información intra e inter poblacional. Una gran mayoría de los métodos actuales para estimar la proporción de sustituciones adaptativas se basan en esta prueba, que a su vez se derivó de la prueba de

---

Hudson-Kreitman-Aguadé (Hudson *et al.* 1987; Tabla 1.2). El MKT compara el número y frecuencia de las mutaciones putativamente funcionales (normalmente codificadoras no sinónimas,  $a$ ) respecto a las posiciones que se suponen selectivamente neutras (como las mutaciones sinónimas o intrónicas,  $s$ ). Se combinan los datos de polimorfismo ( $p$ ) y de divergencia ( $k$ ) de estas clases de secuencias. La quintaesencia del MKT consiste en detectar en la divergencia lo que no se puede detectar en el polimorfismo, las mutaciones adaptativas, las cuales se fijan tan rápido que contribuyen a la divergencia pero no al polimorfismo (Figura 1.4). El test supone que si todas las mutaciones son neutras, la razón  $k_a/k_s$  será teóricamente igual a la razón  $p_a/p_s$ . Si se dan mutaciones polimórficas ligeramente deletéreas ( $p_a$ ) entonces la razón  $p_a/p_s$  será superior a la razón  $k_a/k_s$ , puesto que estas mutaciones pueden segregarse en la población pero no llegarán a fijarse. Además, este incremento de la razón  $p_a/p_s$  se puede también observar si ha habido relajación de la selección sobre sitios previamente deletéreos que se han convertido en neutros, incrementando el polimorfismo pero no la divergencia. Si por el contrario, alguna de las mutaciones putativamente funcionales ( $k_a$ ) ha sido fijada por selección adaptativa, entonces la razón  $k_a/k_s$  debería ser superior a la razón  $p_a/p_s$ , puesto que como se ha comentado las mutaciones adaptativas se fijan tan rápido que contribuyen a la divergencia pero no al polimorfismo. Además, este test permite estimar la proporción de sustituciones putativamente funcionales ( $k_a$ ) que se han fijado por selección adaptativa. Este índice se denomina  $\alpha$  (Charlesworth 1994; Eyre-Walker 2006). Si  $\alpha = 0$ , entonces  $k_a/k_s = p_a/p_s$ . Por tanto, es posible calcular la proporción de sustituciones adaptativas con el estimador  $\hat{\alpha} = 1 - (k_s p_a / k_a p_s)$ .

El MKT es robusto a muchas fuentes de error como las variaciones en la tasa de mutación o efectos demográficos, puesto que al utilizar sitios genómicos adyacentes (por ejemplo sinónimos y no sinónimos) afectan por igual a ambos componentes. Sin embargo, las estimas de la tasa de adaptación del MKT ( $\alpha$ ) pueden estar severamente subestimadas en presencia de alelos ligeramente deletéreos (Figura 1.4, Eyre-Walker 2006, Charlesworth y Eyre-Walker 2008, Fay 2011, Messer y Petrov 2013). Para minimizar el impacto de los alelos ligeramente deletéreos se han propuesto diversos métodos. El más sencillo es suponer que las mutaciones ligeramente deletéreas no segregan por encima de cierta frecuencia en la población y por tanto excluir el polimorfismo que se observe por debajo de ese valor (Fay *et al.* 2001; Charlesworth y Eyre-Walker 2008).



**Figura 1.4.** Esquema que muestra la trayectoria de los alelos neutros y selectivos en una población

(a) La probabilidad de fijación de los alelos neutros (líneas negras) equivale a su frecuencia inicial  $1/2N$  y el tiempo medio que se requiere para que esto ocurra es de  $4N_e$  generaciones. Sin embargo, las mutaciones adaptativas (líneas azul claro) que se fijan, lo hacen de forma mucho más rápida. Por ello, cuando se observa el polimorfismo de una población (línea vertical discontinua), se suele detectar alelos neutros a distintas frecuencias pero no alelos adaptativos. En este ejemplo concreto se observarían 4 sustituciones, dos de ellas adaptativas y dos neutras, pero sólo 2 alelos segregando en la población. Con el MKT obtendríamos que la razón  $k_a/k_s$  excede a la razón  $p_a/p_s$  y por tanto se concluiría que la selección natural ha actuado en este *locus*. (b) Sin embargo, cuando se consideran los alelos ligeramente deletéreos (líneas azul oscuro), que segregan en la población pero no llegan a fijarse, para las mismas sustituciones adaptativas y neutras que en el ejemplo anterior, con el MKT se obtendría que la razón  $k_a/k_s$  es menor que la razón  $p_a/p_s$  y por tanto se concluiría que es un *locus* neutro (b).

Mackay y colaboradores (2012) propusieron una extensión intuitiva y sencilla al MKT. Esta aproximación, además de corregir el efecto de los alelos ligeramente deletéreos, también permite estimar los diferentes regímenes de selección. El método requiere estimar la distribución de frecuencias de los alelos menores (MAF, *Minor Allele Frequency*), es decir, la distribución de frecuencias a las que los alelos raros (en polimorfismos dialélicos) se encuentra en una población dada. El objetivo es separar el componente de las mutaciones ligeramente deletéreas del de las neutras en el polimorfismo putativamente funcional ( $p_a$ );  $p_a = p_{a \text{ neutro}} + p_a$

*ligeramente deletéreo*, donde  $p_a$  es el número de sitios segregantes en la clase de sitios que se supone están sujetos a selección,  $p_{a \text{ neutro}}$  es el número de sitios segregantes neutros y  $p_{a \text{ ligeramente deletéreo}}$  el número de sitios segregantes seleccionados débilmente en contra. Para ello: (i) se asume que las variantes ligeramente deletéreas no segregan por encima de un valor de frecuencia dado que en sus análisis se consideró del 5%. (ii) Del polimorfismo neutro (sinónimo en el caso de que se tome estos sitios como referencia neutra) se estima la fracción de sitios que segregan de forma neutra por debajo del 5% como  $f_{\text{neutra MAF} < 5\%} = p_{s \text{ MAF} < 5\%} / p_s$ . (iii) A partir  $f_{\text{neutra MAF} < 5\%}$  se estima la fracción de variantes neutras que segregan en  $p_a$ :  $p_{a \text{ neutro MAF} < 5\%} = p_a \times f_{\text{neutra MAF} < 5\%}$ . (iv) El resto de los alelos menores al 5% se consideran como alelos ligeramente deletéreas y se calcula como:  $p_{a \text{ ligeramente del}} = p_{a \text{ MAF} < 5\%} - p_{a \text{ neutro MAF} < 5\%}$ . Por tanto, (v) el número esperado de sitios polimórficos neutros en  $p_a$  es;  $p_{a \text{ neutro}} = p_{a \text{ neutro MAF} < 5\%} + p_{a \text{ MAF} > 5\%}$ . Una vez  $p_a$  se ha separado en dos componentes,  $p_{a \text{ ligeramente del}}$  y  $p_{a \text{ neutro}}$ , este método permite estimar los diferentes regímenes de selección que se recogen en la Tabla 1.3 además de obtener estimas corregidas de  $\alpha$  para los alelos ligeramente deletéreas (siempre y cuando estos segreguen mayoritariamente a frecuencias por debajo del 5%).

**Tabla 1.3.** Estimadores de los regímenes de selección derivados de la extensión del MKT

Fracción selectiva	Símbolo	Estimador	Definición
Sitios fuertemente deletéreas	$\hat{d}$	$\hat{d} = 1 - (\hat{b} + \hat{f}) = 1 - \frac{p_a \cdot m_s}{p_s \cdot m_a}$	Fracción de mutaciones nuevas que son fuertemente deletéreas y por lo tanto no segregan
Sitios ligeramente deletéreas	$\hat{b}$	$\hat{b} = \frac{p_{a \text{ ligeramente deletéreas}} \cdot m_s}{p_s \cdot m_a}$	Exceso de sitios que segregan con MAF < 5% respecto al componente neutro
Sitios neutros	$\hat{f}$	$\hat{f} = \frac{p_{a \text{ neutro}} \cdot m_s}{p_s \cdot m_a}$	La fracción neutra estimada del componente neutro tras la corrección de los alelos ligeramente deletéreas
Sitios recientemente neutros	$\hat{\gamma}$	$\hat{\gamma} = \left( \frac{p_{a \text{ neutro}}}{p_s} - \frac{k_a}{k_s} \right) \frac{m_s}{m_a}$	Subconjunto de $f$ . El exceso relativo de polimorfismo respecto a la divergencia una vez eliminada la fracción ligeramente deletérea
Fijaciones adaptativas	$\hat{\alpha}$	$\hat{\alpha} = 1 - \frac{p_{a \text{ neutro}} \cdot k_s}{p_s \cdot k_a}$	Fracción de sustituciones adaptativas tras la corrección de los alelos ligeramente deletéreas

$m$  = longitud de la secuencia. Tabla extraída de Métodos suplementarios 2.6 de Mackay *et al.* 2012.

Desarrollos más sofisticados del MKT intentan inferir la distribución de los efectos de la eficiencia biológica (DFE, *Distribution of Fitness Effect*) de los sitios putativamente funcionales ( $a$ ) a partir del espectro de frecuencias (SFS, *Site Frequency Spectrum*) del polimorfismo en estos sitios para después inferir el número de sustituciones no adaptativas que se esperan (Bustamante *et al.* 2002; Bustamante *et al.* 2005; Eyre-Walker *et al.* 2006; Eyre-Walker y

---

Keightley 2007; Keightley y Eyre-Walker 2007; Boyko *et al.* 2008; Charlesworth y Eyre-Walker 2008; Eyre-Walker y Keightley 2009). Esta aproximación permite corregir las estimas de  $\alpha$ . Además, algunos métodos intentan corregir adicionalmente los posibles efectos de la demografía inferida a partir del SFS de los sitios putativamente neutros ( $s$ ) para después corregir este efecto en el SFS de los sitios putativamente funcionales ( $a$ ) (Williamson 2005; Keightley y Eyre-Walker 2007; Boyko 2008).

### 1.2.3. Estimaciones de la selección adaptativa

La disponibilidad de datos de polimorfismo a gran escala ha posibilitado el escrutinio de los genomas para encontrar evidencias de selección positiva. Estos datos han permitido probar la presencia de arrastres debidos a selección positiva basándose en la reducción local del polimorfismo y el incremento del desequilibrio de ligamiento (Nielsen 2005), y se han publicado muchos estudios de este tipo (ej., Glinka *et al.* 2003; Akey *et al.* 2004; Wright *et al.* 2005; Voight *et al.* 2006; Sabeti *et al.* 2007; Williamson *et al.* 2007). Sin embargo, el solapamiento de las regiones putativamente adaptativas entre los distintos estudios es muy bajo. En humanos, se ha estimado que sólo solapan entre el 8 - 27% de las regiones (Biswas y Akey 2006, Akey 2009). A pesar de la obtención reciente de un millar de genomas humanos para su análisis variacional (*The 1000 Genomes Project Consortium* 2012; Grossman *et al.* 2014), la prevalencia de la selección positiva en el genoma humano sigue siendo muy controvertida (Akey 2009; Hernandez *et al.* 2011).

Los métodos basados en el MKT (1991, secciones anteriores, 1.2.1 y iii)) son los únicos con los que en teoría se pueden obtener estimas no sesgadas de la proporción de cambios causados por la selección adaptativa (Eyre-Walker 2006). En la mayoría de los estudios que se recogen en la Tabla 1.4 se han utilizado ampliaciones y extensiones del test de MK para intentar corregir el sesgo introducido por los alelos ligeramente deletéreos (véase sección iii)). En humanos así como en la mayoría de las plantas y en levaduras (*Saccharomyces cerevisiae* y *Saccharomyces paradoxus*; Liti *et al.* 2009; Elyashiv *et al.* 2010), la proporción de sustituciones adaptativas ( $\alpha$ ) es baja y no supera el 20% en ninguno de los estudios (Tabla 1.4). En especies tan diversas como ratones, conejos y *Drosophila* los valores de las estimas de selección adaptativa son mucho más elevados (Tabla 1.4). El nivel de evolución adaptativa es todavía mayor en bacterias,  $\alpha > 56\%$  (Charlesworth y Eyre-Walker 2006) y algunos virus, con  $\alpha \sim 50\%$  (Williamson *et al.* 2003) y con el 75-85% de los codones con evidencias de selección adaptativa (Nielsen y Yang 2003).

**Tabla 1.4.** Estimaciones de la proporción de la evolución adaptativa ( $\alpha$ ) en diferentes especies y regiones genómicas

Especie analizada	Taxón externo	Región	Fracción de sustituciones adaptativas ( $\alpha$ )	Referencias
<b>Animales</b>				
<i>Homo sapiens</i>	<i>P. troglodytes</i>	Codificadora	0-9%	Mikkelsen <i>et al.</i> 2005
		Codificadora	6%	Bustamante <i>et al.</i> 2005
		Codificadora	10-13%	Gojobori 2007
		Codificadora	~10-20%	Boyko <i>et al.</i> 2008
<i>Mus musculus</i>	<i>Rattus norvegicus</i>	Codificadora	44%	Halligan <i>et al.</i> 2010
<i>Oryctolagus cuniculus</i>	<i>L. granatensis</i>	Codificadora	A ~60% X 52%	Carneiro <i>et al.</i> 2012
<i>D. melanogaster</i>	<i>D. simulans</i>	Codificadora	86%	Andolfatto <i>et al.</i> 2011
		Codificadora	~50%	Eyre-Walker y Keightley 2009
		Codificadora	A ~12,8% X 46%	Langley <i>et al.</i> 2012
		Intrón	~20%	Eyre-Walker y Keightley 2009
	<i>D. yakuba</i>	Genoma	~25%	Mackay <i>et al.</i> 2012
		Codificadora	24%	Mackay <i>et al.</i> 2012
		UTR	7%	Mackay <i>et al.</i> 2012
		Intrón	30%	Mackay <i>et al.</i> 2012
		Intergénica	29%	Mackay <i>et al.</i> 2012
<b>Plantas</b>				
<i>Schiedea globosa</i>	-	Codificadora	~20%	Gossmann <i>et al.</i> 2010
<i>Helianthus petiolaris</i>	-	Codificadora	~50%	Gossmann <i>et al.</i> 2010
<i>Helianthus annuus</i>	-	Codificadora	0%	Gossmann <i>et al.</i> 2010
<i>Populus tremula</i>	-	Codificadora	~10%	Gossmann <i>et al.</i> 2010
<i>Populus balsamifera</i>	-	Codificadora	~5%	Gossmann <i>et al.</i> 2010
<i>Zea mays</i>	-	Codificadora	0%	Gossmann <i>et al.</i> 2010
<i>Capsella grandiflora</i>	<i>N. paniculata</i>	Codificadora	40%	Slotte <i>et al.</i> 2010
<i>Arabidopsis thaliana</i>	<i>A. lyrata</i>	Codificadora	0%	Slotte <i>et al.</i> 2010



<b>Bacterias</b>				
<i>Escherichia coli</i>	<i>Salmonella enterica</i>	Codificadora	>56%	Charlesworth y Eyre-Walker 2006
<b>Virus</b>				
<i>VIH</i>	-	Codificadora	50%	Williamson 2003

A = autosomas, X = cromosoma X.

Estas diferencias en los niveles de evolución adaptativa pueden indicar bien que el MKT y sus variantes están sesgados en diferente grado y dirección en estos linajes (Fay *et al.* 2011) o simplemente reflejar la variación real en la tasa de adaptación en diferentes especies. Se ha sugerido que las grandes diferencias observadas entre linajes e incluso entre especies filogenéticamente cercanas pueden ser debidas a la variación en el censo efectivo de la población (Gossmann *et al.* 2012; Romiguier *et al.* 2014). Aunque exista cierta controversia sobre la verdadera relevancia de la selección adaptativa en ciertas especies, parece que la selección adaptativa juega un papel más relevante que el atribuido por la teoría neutralista (Tabla 1.4) al menos en *Drosophila melanogaster*, especie focal de la investigación de esta tesis.

#### 1.2.4. Estimaciones de la selección negativa

En las regiones codificadoras (los sitios no sinónimos) de diversas especies, la selección negativa parece ser ubicua (Tabla 1.5). Por otra parte, la detección de selección negativa fuerte en regiones no codificadoras ayuda a comprender cómo evolucionan dichas regiones. Al comparar la divergencia entre humanos y ratones se estimó que 5% del genoma está sujeto a selección purificadora (Waterston 2002). Estudios posteriores han incrementado esta proporción al 9% del genoma humano (Smith *et al.* 2004). Suponiendo que 1,5% del genoma humano codifica para proteínas (Venter *et al.* 2001), el que un 9% del genoma esté sometido a selección purificadora implica que el 83% de las secuencias constreñidas corresponden a regiones no codificadoras. Estas estimaciones son superiores en *Drosophila* (Tabla 1.5). En estudios iniciales a partir de información sobre cientos de genes individuales se estimó que el 40% - 60% de las regiones no codificadoras están muy constreñidas. Estudios más recientes, con genomas completos de decenas de individuos, calculan que ~35% de los sitios UTR, ~32% de los intrones y el ~40% de las regiones intergénicas evolucionan bajo selección negativa fuerte (Mackay *et al.* 2012).

La selección purificadora puede también ser débil permitiendo que los alelos segreguen en la población aunque rara vez se fijen. Estos alelos ligeramente deletéreos parecen ser muy comunes en distintas poblaciones naturales debido a que se ha detectado un exceso de alelos polimórficos a baja frecuencia para los sitios putativamente funcionales, como los codificadores

(Bustamante *et al.* 2005; Lohmueller 2011; Eyre-Walker y Keightley 2007; Mackay *et al.* 2012).

**Tabla 1.5.** Estima del porcentaje de alelos bajo selección negativa en distintas especies y regiones funcionales

<b>a) Regiones codificadoras en distintas especies</b>		
<b>Especie</b>	<b>Porcentaje de alelos bajo selección negativa</b>	<b>Referencias</b>
<i>Homo sapiens</i>	~77% ~43% ~67%	Mikkelsen <i>et al.</i> 2005 Boyko <i>et al.</i> 2008 Eyre-Walker y Keightley 2009
<i>Mus musculus</i>	~87%	Mikkelsen <i>et al.</i> 2005
<i>D. melanogaster</i>	~86% ~78%	Halligan y Keightley 2006 Mackay <i>et al.</i> 2012
<i>Capsella grandiflora</i>	~86%	Slotte 2010
<i>Arabidopsis thaliana</i>	~66%	Slotte 2010

<b>b) Regiones no codificadoras en D. melanogaster</b>		
<b>Regiones funcionales</b>	<b>Porcentaje de alelos bajo selección negativa</b>	<b>Referencias</b>
<i>UTR</i>	~60% ~35%	Andolfatto 2005 Mackay <i>et al.</i> 2012
<i>Intrón</i>	~40% ~53% ~32%	Andolfatto 2005 Halligan y Keightley 2006 Mackay <i>et al.</i> 2012
<i>Integénico</i>	~44% ~50% ~56% ~40%	Halligan <i>et al.</i> 2004 Andolfatto 2005 Halligan y Keightley 2006 Mackay <i>et al.</i> 2012

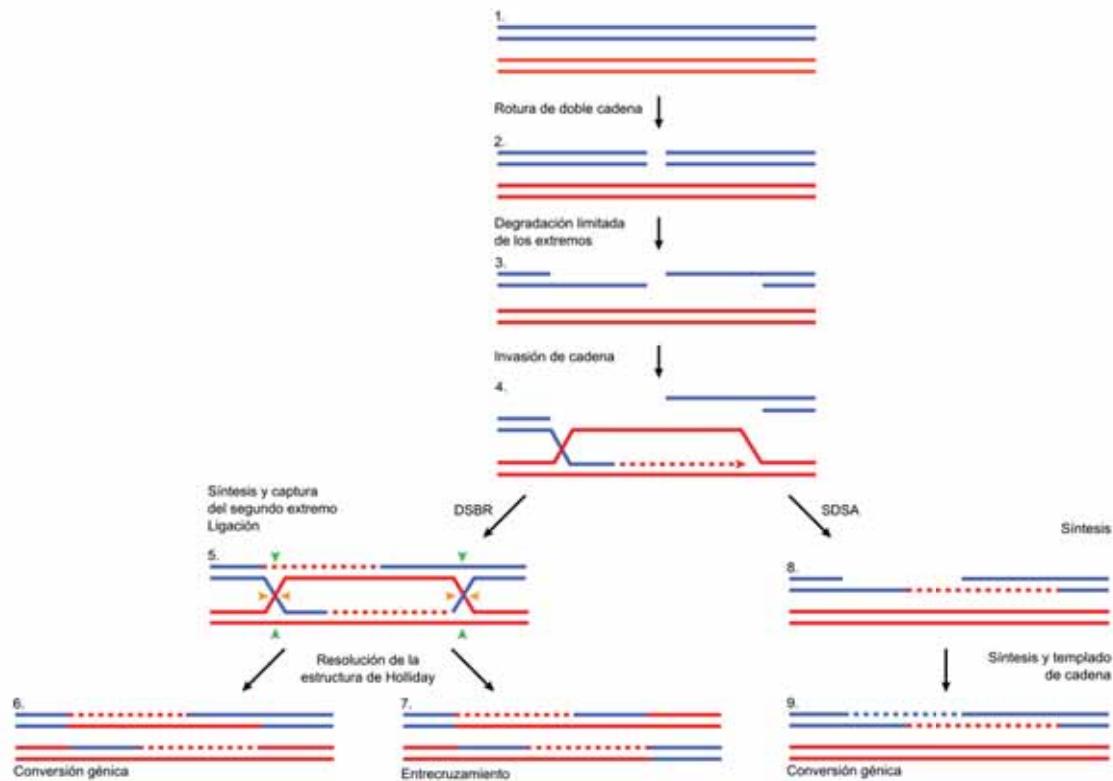
### 1.3. Recombinación en el nivel molecular

La recombinación homóloga es un proceso mediante el cual un par de secuencias homólogas de DNA intercambian una parte de su secuencia. Es un proceso que normalmente ocurre entre dos copias homólogas del mismo cromosoma aunque también puede darse entre moléculas no homólogas similares de DNA. Los conocimientos sobre los mecanismos de recombinación fueron inicialmente estudiados en bacterias y levaduras (revisados en Keeney 2001, Baudat y Keeney 2001, Lichten 2001). Debido a su corto tiempo de generación y al tamaño reducido de sus genomas, se pudieron aislar cepas mutantes para procesos recombinantes defectuosos y caracterizar así las proteínas asociadas a este proceso. Más tarde, se descubrieron y caracterizaron proteínas homólogas en *Drosophila*, ratones y humanos (Lichten 2001).

---

El proceso de recombinación generalmente aceptado en eucariotas comienza con la rotura de doble cadena en uno de los cromosomas homólogos (Sung y Klein 2006). Tras una serie de procesos intermedios (Figura 1.5) culmina en dos posibles resultados: la conversión génica y/o el entrecruzamiento. En la conversión génica, sólo se intercambia una pequeña parte de la información génica entre cromosomas. En humanos, se ha estimado que el 90% de las roturas de doble cadena se resuelven por conversión génica (Jeffreys y May 2004). Sin embargo, la cantidad de material intercambiado entre los cromosomas (lo que se conoce como tractos de conversión) se reduce a ~300pb (Jeffreys y May 2004). En *Drosophila*, se ha estimado que la razón de las roturas de doble cadena que se resuelven por entrecruzamiento respecto a los que se resuelven por conversión génica es aproximadamente 4 (Hilliker y Chovnick 1981; Hilliker *et al.* 1991). A su vez, se han efectuado varias estimas de la longitud de los tractos en *loci* específicos (Hilliker *et al.* 1994; Betrán *et al.* 1997, Blanton *et al.* 2005) y en el cromosoma X (Miller *et al.* 2012), siendo en promedio ~340pb.

Un estudio reciente basado en datos genómicos de varios individuos (Cuadro 3, sección 3.4 de Materiales y métodos; Comeron *et al.* 2012) calcula que aproximadamente entre el ~83,0% de los sucesos de recombinación se resuelven por conversión génica, es decir que la razón entrecruzamiento *versus* conversión génica es ligeramente superior a la estimada previamente (~4). La longitud promedio de los tractos de conversión promedio fue de 518pb (Comeron *et al.* 2012), un poco mayor que la estimada para *loci* específicos (352pb en Hilliker *et al.* 1994; 122pb en Betrán *et al.* 1997, 441pb en Blanton *et al.* 2005) o el cromosoma X (476pb; Miller *et al.* 2012). En consecuencia, a pesar de que la conversión génica parece ser el resultado más frecuente del intercambio entre cromátidas homólogas no hermanas, el rastro que deja en la secuencia de DNA es tan corto (unos cientos de pb) que su detección precisa sigue siendo un reto hoy día (Hellenthal y Stephens 2006; Gay y McVean 2007). La resolución alternativa, el entrecruzamiento, intercambia grandes fragmentos de material genético por lo que su efecto se extiende mucho más en el genoma. Excepto que haya un doble entrecruzamiento entre las mismas cromátidas, durante el entrecruzamiento se intercambia todo el material genético desde el punto de intercambio hasta el final del cromosoma, lo que hace que sea más fácil de detectar tanto experimentalmente como mediante el análisis bioinformático de secuencias.



**Figura 1.5.** Modelo molecular de recombinación homóloga en eucariotas

El proceso empieza con una rotura de doble cadena (2) en uno de los cromosomas. Las regiones adyacentes a la rotura de doble cadena se degradan mediante una exonucleasa, lo que genera pequeños fragmentos de DNA de cadena simple que puede invadir y unirse al DNA homólogo del otro cromosoma el cual se usa como molde para la reparación (3,4). Se conocen dos mecanismos distintos a partir de este punto: (a) El mecanismo DSBR (*double-strand break repair*) en el que se forma la estructura de Holliday (Holliday 1964) (5). Esta estructura inestable puede resolverse de dos maneras: la conversión génica (6) o el entrecruzamiento (7). (b) Alternativamente, las roturas de doble cadena pueden también resolverse mediante el mecanismo de templado de cadena dependiente de síntesis o SDSA (*synthesis-dependent strand annealing*) entre cromátidas no hermanas (8). Esta vía se resuelve mediante la conversión génica (9). Figura adaptada de Sung y Klein (2006).

La mayoría de los modelos de la genética de poblaciones que tienen en cuenta la recombinación la consideran una fuerza que reorganiza la relación entre *loci* y no dentro de los *loci*. La recombinación intragénica o la conversión génica se ha considerado irrelevante históricamente puesto que el entrecruzamiento, al intercambiar fragmentos promedio mucho mayores que la conversión, se veía como el factor casi exclusivo de recombinación (Berry y Barbadilla 2000). Sin embargo, tras la observación del impacto de los arrastre por selección positiva y los arrastres por selección negativa a nivel genómico (Begun y Aquadro 1991; Charlesworth 1993; Betancourt y Presgraves 2002; Presgraves 2005; Begun *et al.* 2007; Betancourt *et al.* 2009; Sella *et al.* 2009; Mackay *et al.* 2012; Langley *et al.* 2012), no cabe duda del papel central que juega la recombinación en la genética de poblaciones moderna y la conversión génica se ha convertido en un factor muy a tener en cuenta. La importancia relativa de la conversión génica y el entrecruzamiento depende completamente de la distancia entre

sitios (Berry y Barbadilla 2000). Para regiones de DNA distantes en relación a la longitud de intercambio, lo que en la mayoría de los casos supone distancias intergénicas, el entrecruzamiento es la forma de recombinación predominante (Hilliker *et al.* 1994; Navarro *et al.* 1996). Sin embargo, a escala local, intragénica, la conversión génica parece ser la forma de recombinación mayoritaria (Berry y Barbadilla 2000). Durante los últimos años, con el inicio de la genómica de poblaciones, el entrecruzamiento se ha analizado y estudiado desde un punto de vista genético poblacional (Smukowski y Noor 2011; Cutter y Payseur 2013), aunque el análisis de la conversión génica continúa siendo difícil (Hellenthal y Stephens 2006; Gay *et al.* 2007) y su fenomenología junto con las consecuencias que conlleva son poco conocidas.

**Tabla 1.6.** *Estimas de la tasa de recombinación en distintas especies*

Especie	cM/Mb	Referencias
<b>Animales</b>		
<i>Homo sapiens</i>	1,19	Kong <i>et al.</i> 2002 Ptak <i>et al.</i> 2004
<i>Rattus norvegicus</i>	0,6	Jensen-Seaman <i>et al.</i> 2004
<i>Mus musculus</i>	0,56	Jensen-Seaman <i>et al.</i> 2004
<i>Drosophila</i>	2,31	Fiston-Lavier <i>et al.</i> 2010 Comeron <i>et al.</i> 2012
<i>Caenorhabditis elegans</i>	3,06	Barnes <i>et al.</i> 1995 Cutter 2006
<b>Plantas</b>		
<i>Zea Mays</i>	0,7	Fu <i>et al.</i> 2002 Tenaillon <i>et al.</i> 2004
<i>Arabidopsis thaliana</i>	4,6	Mezard 2006 Wright <i>et al.</i> 2003b
<b>Levaduras</b>		
<i>Saccharomyces cerevisiae</i>	370	Petes 2001
<b>Virus</b>		
<i>VIH</i>	30.000	Jetzt <i>et al.</i> 2000

La tasa de recombinación se suele medir calculando el número de sucesos de recombinación entre marcadores. Así, se puede obtener una medida de la distancia entre *loci* expresado en centiMorgans (*cM*). A lo largo de los años se han construido mapas genéticos para diversas especies (Tabla 1.6). En *Drosophila* la tasa de recombinación promedio por *cM*, calculada a partir de mapas genéticos clásicos, es de 2,295 *cM/Mb* (Fiston-Lavier *et al.* 2010). A pesar de que las diferencias de los promedios entre brazos cromosómicos no son muy grandes, parece haber una gran heterogeneidad dentro de los cromosomas. Hay una reducción importante de la recombinación en las regiones centroméricas y una abrupta reducción en el

extremo de las regiones teloméricas. Durante el desarrollo de esta tesis, se publicó un mapa de recombinación con un número muy superior de marcadores y por tanto con una resolución genética mucho mayor que el mapa basado en marcadores clásicos (Cuadro 3; Comeron *et al.* 2012). Aunque la variabilidad y la resolución de la recombinación es mayor y por lo tanto más precisa dentro de los cromosomas, el patrón general de recombinación sigue siendo muy similar al previamente descrito (valor de recombinación promedio del genoma = 2,325 cM/Mb).

Una alternativa a los métodos experimentales previamente mencionados para la estimación de la tasa de recombinación, son los basados en el análisis estadístico de los datos de variación genética de las poblaciones. En este caso se estima el parámetro de recombinación ( $4N_e r$ ) en lugar de la tasa de recombinación ( $r$ ). El desequilibrio de ligamiento (LD) entre alelos o sitios polimórficos próximos en una región cromosómica hace que un alelo en una posición sea informativo del alelo que se encuentra en otro *locus* ligado. Como los sucesos de recombinación rompen los bloques de ligamiento entre *loci*, se podrían usar los estimadores tradicionales de LD (Tabla 1.7) como medida indirecta de la tasa de recombinación. Sin embargo, el LD está influenciado por otros factores además de la recombinación como son la mutación, la deriva genética, el grado de consanguinidad y la selección natural (Nei 1987). Todos estos efectos pueden enmascarar el impacto de la recombinación sobre el LD, haciendo la correlación entre estos dos factores sea baja, inexistente o engañosa (Litt y Jorde 1986; Weir 1989; Hill y Weir 1994; Thompson *et al.* 1988, Walter y Cox 1991; Devlin y Risch 1995). Por eso, se han buscado estimadores más sofisticados que permitan determinar la tasa de recombinación a partir de los patrones de LD.

**Tabla 1.7.** *Estimadores del desequilibrio de ligamiento (LD)*

Estadístico	Descripción	Referencias
$D'$	Medida de desequilibrio de ligamiento definida como la diferencia entre la frecuencia de un haplotipo digénico (o dos <i>loci</i> ) y el producto de las frecuencias de los alelos que lo componen, dividido por el valor más extremo posible de desequilibrio dadas las frecuencias alélicas marginales	Lewontin (1964)
$r^2$	Correlación estadística (cuadrado del coeficiente de asociación) de las frecuencias génicas entre dos <i>loci</i> ligados	Hill y Robertson (1968)

Los estimadores más recientes utilizan la teoría de la coalescencia (Cuadro 2) para corregir el efecto de diversos factores sobre el LD y así poder generar un modelo de sucesos de recombinación. A partir de los datos de variación nucleotídica se utilizan métodos de máxima verosimilitud para inferir el parámetro de recombinación ( $4N_e r$ ) según estos modelos. Calcular los valores de máxima verosimilitud en secuencias genómicas bajo un modelo de coalescencia es computacionalmente prohibitivo, excepto para un conjunto de datos mucho más reducido

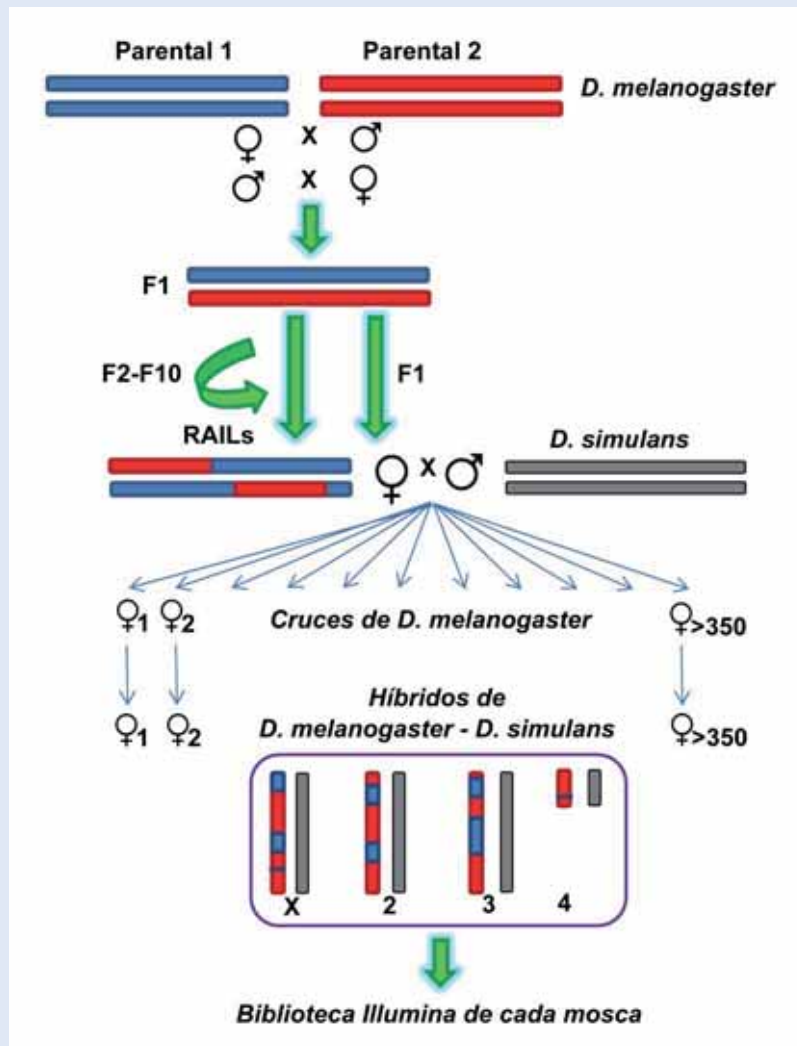
---

(Fearnhead y Donnelly 2001). Por tanto, para obtener estimas del parámetro de recombinación en cromosomas o genomas completos, los datos se pueden analizar por ventanas o regiones más pequeñas. En este caso, la máxima verosimilitud de los datos se estima para cada ventana o región y posteriormente estas estimas se combinan en un valor de máxima verosimilitud compuesto (*composite likelihood*). Este método fue inicialmente propuesto por Hudson (2001), y repetidamente ampliado para incluir modelos de mutación compleja (McVean *et al.* 2002) y estimas variables de tasas de recombinación (McVean *et al.* 2004). Las últimas versiones del método son suficientemente rápidas como para aplicarlos en el nivel de genomas completos. De hecho, este método de aproximación de máxima verosimilitud implementado en el software *Ldhat* (McVean *et al.* 2004) se ha utilizado ampliamente en los últimos años en el campo de la genómica y genética de poblaciones en diferentes grupos de organismos (*Humanos*: Laayouni *et al.* 2011; Manku *et al.* 2013; Alves *et al.* 2014; *Ratón*: Brunshwing *et al.* 2012; *Vaca*: Sodeland *et al.* 2011; *Perro*: Vaysse *et al.* 2011, Axelsson *et al.* 2012; *Jabalí*: Bedada *et al.* 2014; *Drosophila*: Langley *et al.* 2012; *Levadura*: Tsai *et al.* 2010, *Aves*: Huynh *et al.* 2011, *Peces*: Feulner *et al.* 2015; *Plantas*: Paape *et al.* 2012; Morris *et al.* 2012; Choi *et al.* 2013).

Las diferencias principales entre los métodos de recombinación que estiman la tasa de recombinación directamente mediante el análisis de recombinantes en individuos ( $r$ ) y los métodos basados en el análisis de variación genómica para la estimación del parámetro de recombinación ( $4N_e r$ ) son: (i) el parámetro de recombinación implícitamente recoge información histórica de la población (la contenida en la variación nucleotídica) mientras que la tasa de recombinación es la tasa actual (que ha ocurrido en pocas generaciones) de la población. (ii) El parámetro de recombinación se puede estimar para diferentes tamaños de ventana a lo largo del genoma mientras que la tasa de recombinación está disponibles para una resolución y tamaño de ventana determinado y por tanto no se pueden obtener estimas fiables a ventanas menores. Además, (iii) las limitaciones asociadas a uno u otro método difieren. Así, el parámetro poblacional puede tener un exceso de falsos positivos debido a que otras fuerzas evolutivas pueden imitar el patrón que deja la recombinación en las secuencias genómicas mientras que en la estimación de la tasa de recombinación no se tiene en cuenta la variación histórica de la recombinación a lo largo del genoma. Por lo tanto, el uso de ambos métodos permite obtener estimas de recombinación complementarias sobre la recombinación en una especie concreta.

**Cuadro 3. Cartografía genómica de la recombinación a alta resolución en *D. melanogaster*, c**

Comeron y colaboradores (2012) han estimado en un trabajo pionero la tasa de recombinación a alta resolución en *D. melanogaster*, distinguiendo entre el entrecruzamiento y la conversión génica, en el nivel genómico y para distintos individuos. Para ello, se llevaron a cabo 8 cruces recíprocos entre 10 líneas del proyecto DGRP (sección 1.6), una línea de Madagascar y otra de Papúa Nueva Guinea. De los descendientes de estos cruces (F<sub>1</sub>) se realizaron cruces hermano-hermana para generar las líneas homocigóticas denominadas RAIL (*Recombinant Advanced Intercross Lines*) (Figura 1.6). Las hembras RAIL se cruzaron con machos *D. simulans*. Finalmente, las hembras híbridas de *D. melanogaster* se secuenciaron. Se eliminaron las lecturas potenciales de *D. simulans* y se seleccionaron aquellas que cartografiaban únicamente en uno de las dos líneas parentales. Así, se pudieron obtener los haplotipos (potencialmente recombinantes) de cada brazo cromosómico de las distintas hembras híbridas de cada cruce (Figura 1.6).

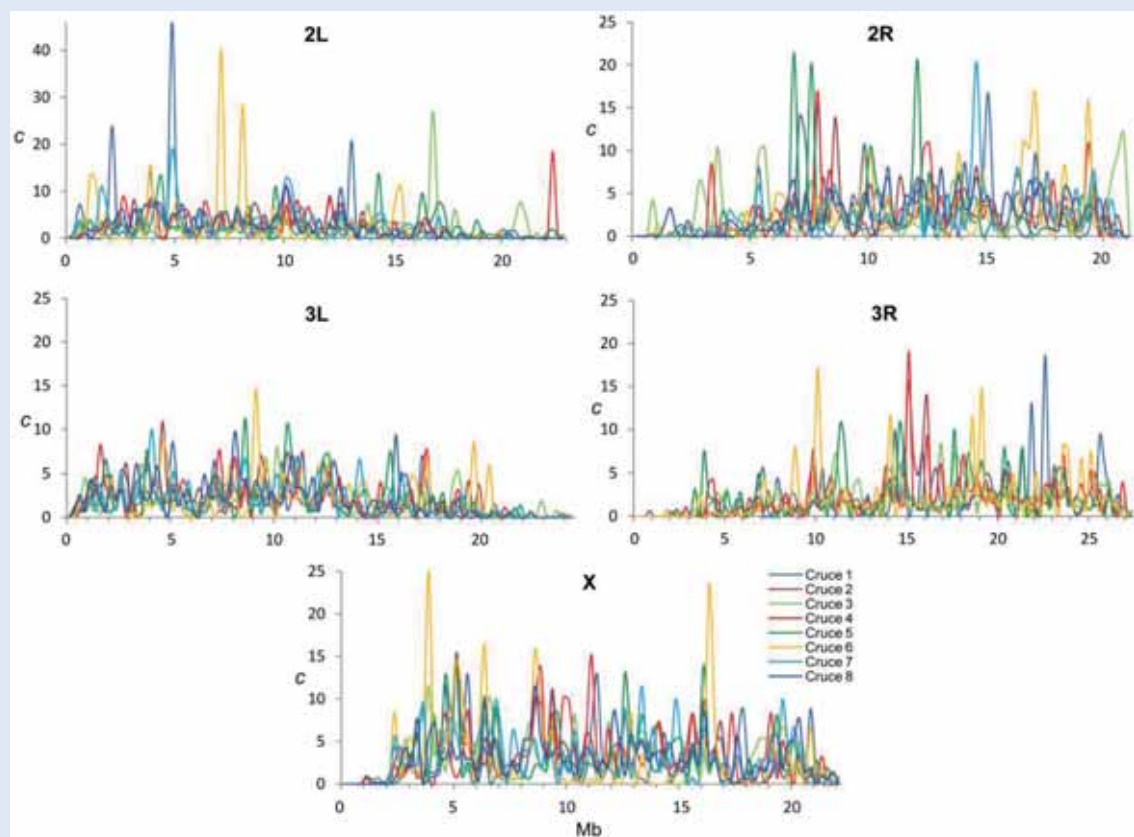


**Figura 1.6.** Diseño experimental seguido para obtener las estimas de recombinación

Representación esquemática de los cruces llevados a cabo para generar cromosomas recombinantes entre las líneas parentales de *D. melanogaster*. Figura tomada de Comeron *et al.* 2012.



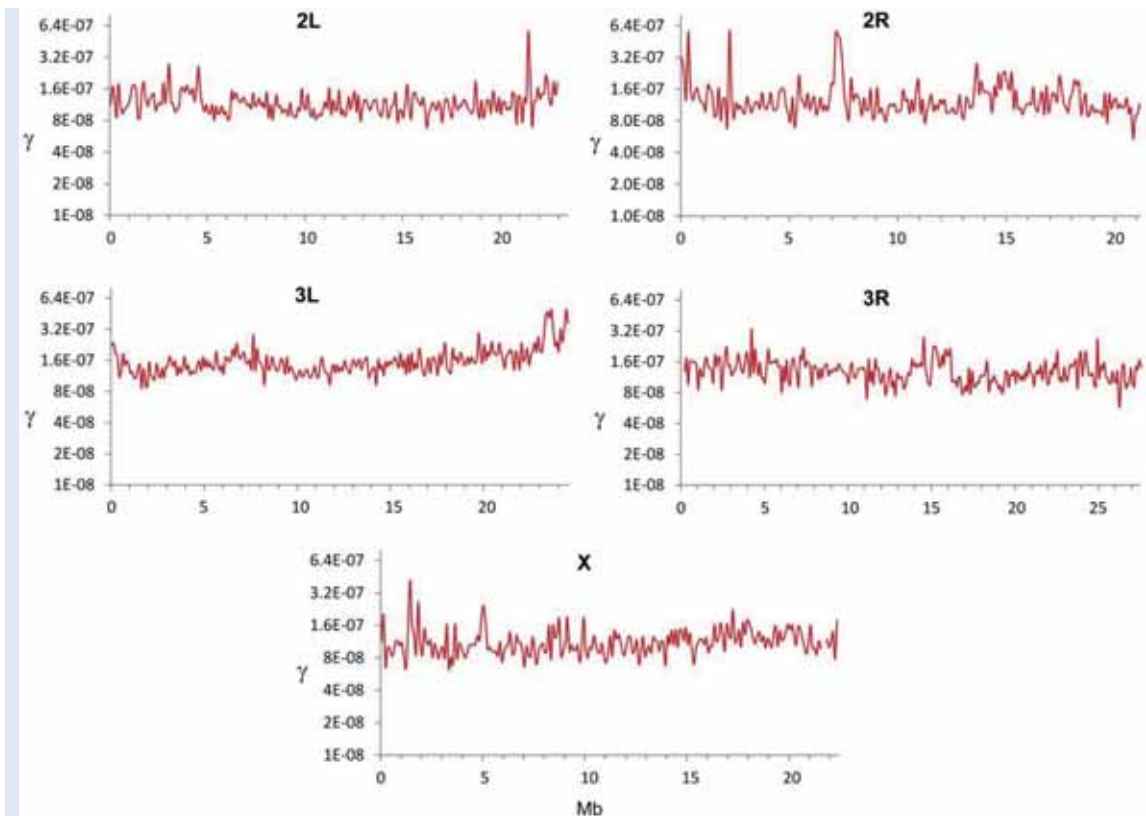
La tasa de entrecruzamiento muestra el patrón general previamente conocido con una reducción paulatina desde el inicio del tercio correspondiente a la región centromérica avanzando en la dirección hacia el centrómero y una drástica reducción en la proximidad del telómero. Sin embargo, se observan que en las regiones previamente consideradas de alta recombinación (distintas de las centroméricas o teloméricas) las tasas de recombinación varían entre 15 y 20 veces entre ellas. Es decir, que hay una gran heterogeneidad intracromosómica para la tasa de entrecruzamiento. Además, la tasa de entrecruzamiento también muestra heterogeneidad entre individuos (Figura 1.7). La correlación en la tasa de entrecruzamiento entre individuos de los 8 cruces es significativo pero más bien débil ( $R$  de Spearman = 0,451). Los cromosomas 2 y 3R son los más diversos, mientras que el X y principalmente el 3L son los más homogéneos entre los distintos individuos. Cuando se combinan los 8 mapas, parece que las regiones más variables en la tasa de recombinación están fuertemente influenciadas por puntos calientes (*hotspot*) de recombinación a baja frecuencia.



**Figura 1.7.** Tasa de recombinación por brazo cromosómico y cruce

Variación individual en la tasa de entrecruzamiento a lo largo de los brazos cromosómicos, de cada uno de los ocho cruces efectuados (cada uno de un color) en ventanas no solapantes de 250kb. Figura tomada de Comeron *et al.* 2012.

La longitud del tracto de la conversión génica es de 518pb en promedio del genoma y el ~83% de los eventos iniciales de recombinación homóloga se resuelven como conversión génica. La distribución de la tasa de conversión génica a lo largo de los cromosomas es mucho más homogénea que la tasa de entrecruzamiento, sin aparente disminución en las regiones centroméricas y teloméricas (Figura 1.8).



**Figura 1.8.** Tasa de conversión génica a lo largo de los brazos cromosómicos

Estimas de la tasa de inicio de la conversión génica ( $\gamma$ , número de eventos/pb  $\cdot$  meiosis de hembra) a lo largo de los brazos cromosómicos en ventanas adyacentes de 100kb. Figura tomada de Comeron *et al.* 2012.

Hay una correlación negativa entre la tasa de entrecruzamiento y la tasa de inicio de la conversión génica ( $R$  de Spearman = -0,1246; valor- $p$  = 0) incluso si no se tienen en cuenta las regiones teloméricas y centroméricas ( $R$  de Spearman = -0,1991; valor- $p$  = 0,0001). Esto probablemente sea debido a un uso diferente de las dos vías de reparación de las roturas de doble cadena (DSBR o SDSA, Figura 1.5) o una resolución diferencial de la estructura de Holliday a lo largo del genoma (Figura 1.5). Esta correlación negativa también se observa entre el polimorfismo ( $\pi$ ) y la tasa de inicio de la conversión génica ( $R$  = -0,560, valor- $p$  = 0; sin regiones centro/teloméricas:  $R$  = -0,499, valor- $p$  = 0) o longitud del tracto (-0,197, valor- $p$  = 0, sin regiones centro/teloméricas:  $R$  = -0,175, valor- $p$  = 0).

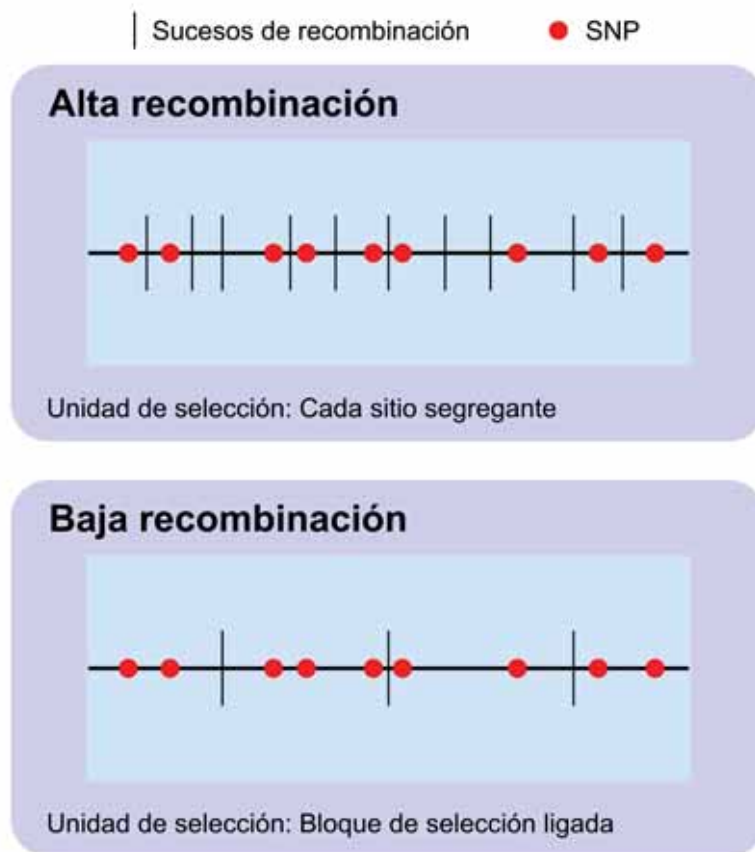
## 1.4. Retos a la teoría neutralista

### 1.4.1. La selección ligada y el efecto de la recombinación

Evidencias empíricas en los patrones de variación nucleotídica y de divergencia en los genomas de diferentes especies descritas hasta el momento (Betancourt y Presgraves 2002; Presgraves 2005; Begun *et al.* 2007; Betancourt *et al.* 2009) junto a estudios teóricos y de simulación recientes llevados a cabo en el ámbito de la genética de poblaciones (Comeron y Kreitman 2002; Comeron *et al.* 2008; Hernandez 2008; Charlesworth 2009; Messer y Petrov 2013) han desplazado el eterno debate neutralista-seleccionista (Lewontin 1974) hacia una

nueva perspectiva: la recombinación es la fuerza evolutiva clave que determina la importancia relativa de la deriva genética *versus* la selección natural adaptativa en el nivel del DNA.

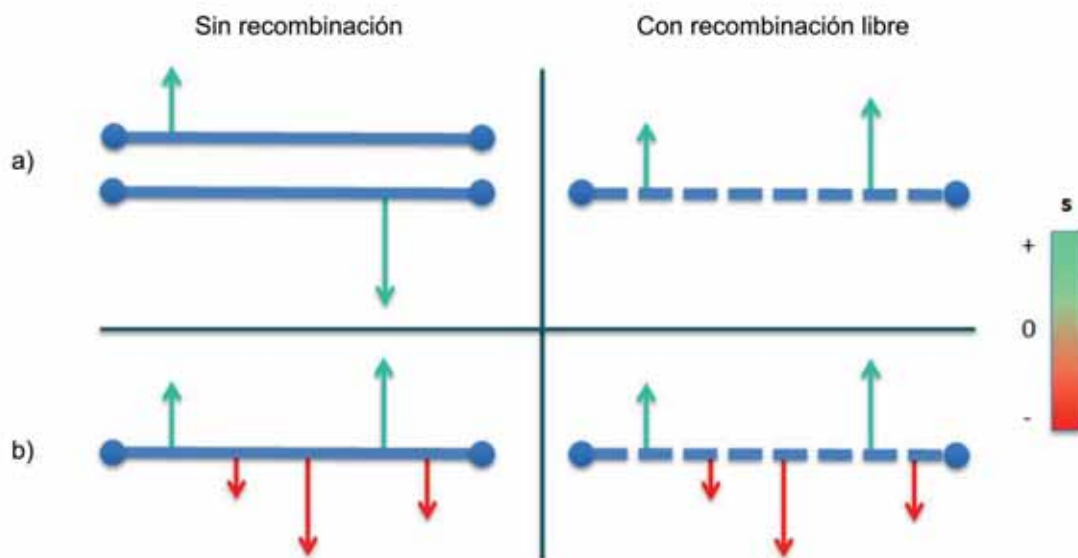
Si la tasa de entrecruzamiento es suficientemente elevada como para hacer que los sitios polimórficos sean independientes unos de otros, la evolución molecular se puede comprender como el resultado de las diversas fuerzas evolutivas actuando sobre cada uno de los sitios polimórficos (Figura 1.9). Una gran mayoría de los desarrollos teóricos de la genética de poblaciones, incluyendo la teoría neutralista, se ha llevado a cabo bajo este supuesto de evolución de un único sitio. Por lo tanto, cuando este supuesto se cumple, se puede aplicar de forma precisa el marco teórico de sitio único de la genética de poblaciones al uso (Charlesworth y Charlesworth 2009). Sin embargo, cuando la tasa de entrecruzamiento es baja respecto a la tasa de mutación o la intensidad de la selección, el destino de una nueva mutación se determina en función del efecto colectivo de las fuerzas de selección sobre todos los sitios ligados (Figura 1.9). La unidad de selección no es por lo tanto un sitio polimórfico o SNP, si no todo el bloque sujeto a selección ligada.



**Figura 1.9.** La unidad de selección depende de la tasa de recombinación local

El esquema muestra que en un contexto genómico con una alta tasa de recombinación (arriba) la unidad de selección es cada sitio segregante, mientras que en regiones con baja tasa de recombinación (abajo) la unidad de selección es el bloque de selección ligada. Alta recombinación debe entenderse como aquella tasa de recombinación que frente a la tasa de mutación y la intensidad de selección permite que los sitios adyacentes segreguen de modo efectivamente independiente.

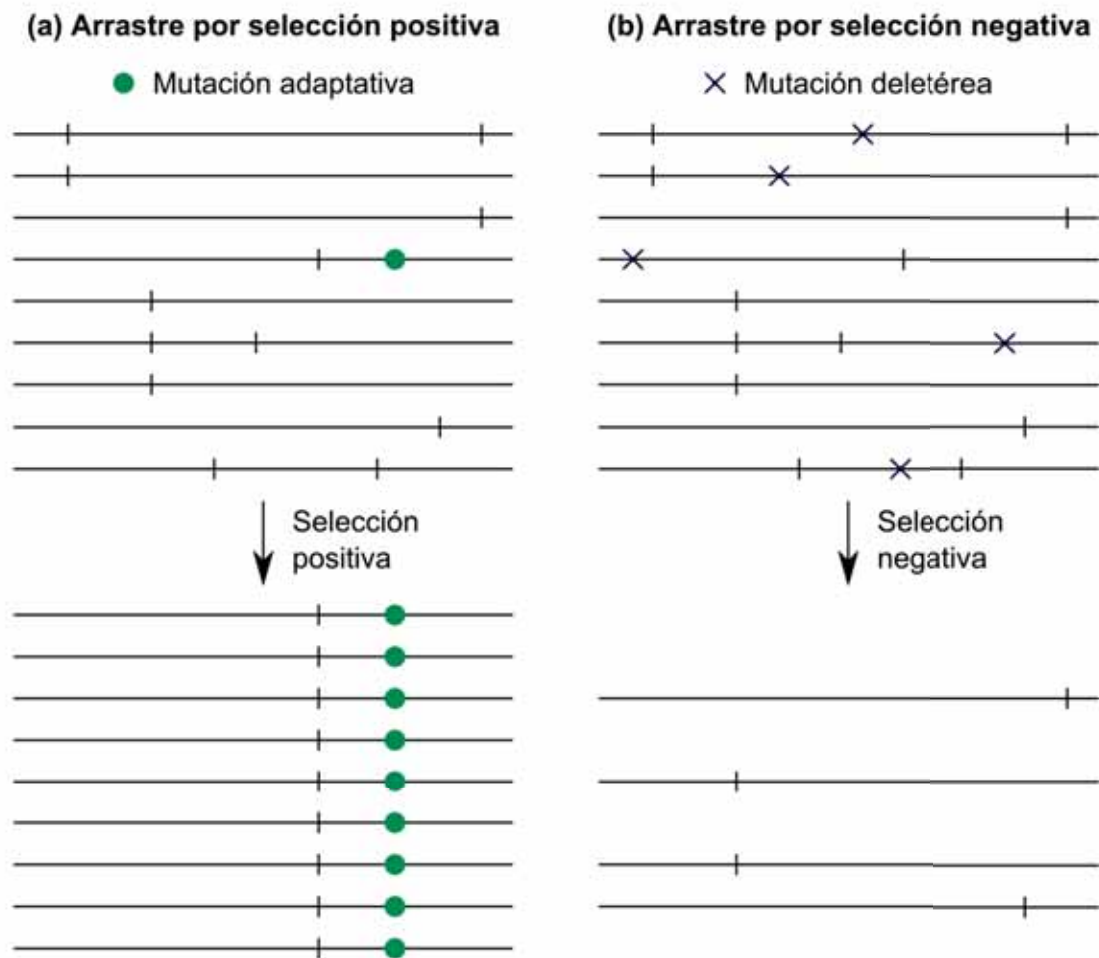
Cuando varios sitios ligados cosegregan en una población finita, su conducta puede interpretarse mediante la interferencia de Hill-Robertson (iHR, Hill y Robertson 1966, Figura 1.10). Si los sitios están sometidos mutuamente a selección, se pueden producir dos escenarios distintos: (i) dos o más mutaciones adaptativas aparecen en haplotipos diferentes en una región con una tasa muy reducida de entrecruzamiento y compiten entre sí por la fijación. Sólo uno de los haplotipos se fijará y el otro se perderá en la población, disminuyendo así la tasa de fijación adaptativa (Figura 1.10a). Por otro lado, (ii) en el mismo haplotipo pueden coexistir mutaciones deletéreas y otras adaptativas. La falta de recombinación forzarán a que en algunos casos mutaciones deletéreas se fijen arrastradas por variantes adaptativas ligadas mientras que en otros casos mutaciones adaptativas se eliminarán arrastradas por las variantes deletéreas ligadas (Figura 1.10b). En conjunto, las consecuencias de la acción de la selección en más de un sitio simultáneamente que se encuentran mutuamente ligados, es que la eficacia de la selección se verá reducida, disminuyendo la tasa de fijación adaptativa e incrementando la fijación de alelos ligeramente deletéreos. Cuanto menor sea la recombinación, mayor será el número de sitios selectivos cosegregando y cuanto mayor sea la intensidad de la selección, mayor será el efecto de la reducción de la eficacia de la selección por la iHR (Comeron *et al.* 2008; Messer y Petrov 2013).



**Figura 1.10.** *Interferencia de Hill-Robertson*

Representación del efecto de la iHR sobre sitios selectivos en un cromosoma dado. Las flechas indican mutaciones adaptativas (verdes) y deletéreas (rojas). La longitud de las flechas indica la intensidad de la selección. (a) Dos o más mutaciones adaptativas. Sin recombinación, sólo una de las mutaciones puede fijarse, es decir, los cromosomas compiten entre sí por la fijación. Si la recombinación es libre ambas mutaciones podrán fijarse de forma rápida. (b) Mutaciones adaptativas y deletéreas. Sin recombinación, los alelos adaptativos y deletéreos competirán entre sí pudiendo fijarse alelos deletéreos cuando la fuerza de selección del alelo adaptativo es fuerte o los alelos adaptativos podrán perderse si la fuerza de selección conjunta de los negativos es superior. Con recombinación, los alelos deletéreos podrán eliminarse de la población y los adaptativos fijarse sin interferir unos con otros.

Los análisis de las tasas de evolución proteica entre especies de *Drosophila* muestran que los genes localizados en regiones genómicas con una reducción importante de la tasa de recombinación tienen un exceso de mutaciones deletéreas y un déficit de fijaciones adaptativas comparado con regiones donde la tasa de recombinación es alta (Hilton *et al.* 1994; Takano 1998; Comeron y Kreitman 2000; Betancourt y Presgraves 2002; Zhang y Parsch 2005; Haddrill *et al.* 2007; Shapiro *et al.* 2007; Campos *et al.* 2014). Esto apoyaría el papel importante que la iHR juega en los genomas y la necesidad de introducir al marco teórico actual de la genética de poblaciones el efecto de los sitios ligados.



**Figura 1.11.** Modelos de arrastre de la variación genética por selección

Esquema de los arrastres por selección positiva y selección negativa sin recombinación. Bajo el modelo de arrastre por selección positiva (a), una mutación adaptativa (círculo verde), se fija rápidamente en la población por selección adaptativa, arrastrando consigo las mutaciones neutras adyacentes (líneas verticales). Tras un arrastre completo por selección positiva y sin recombinación todos los individuos de la población poseerán el mismo haplotipo. Bajo el modelo de arrastre por selección negativa (b), las mutaciones deletéreas (cruces azules) aparecen en diferente haplotipos de la población. Después se irán eliminando arrastrando consigo las variantes neutras ligadas en cada uno de los haplotipos. Figura adaptada de Nachman (2001).

Además, la iHR también afecta a los sitios neutros que están ligados a los sitios selectivos (Comeron *et al.* 2008). Los efectos que sufren los sitios neutros adyacentes debido a la iHR se

---

explican en los modelos de arrastre por selección positiva (Smith y Haigh 1974; Kaplan *et al.* 1989, Berry *et al.* 1991; Gillespie 2000) y de arrastre por selección negativa (Charlesworth *et al.* 1993) (Figura 1.11). De hecho, los modelos de arrastre y la iHR pueden verse como un caso particular del modelo general de selección ligada (Charlesworth 2009).

Bajo la teoría neutralista, suponiendo que los sitios son selectivamente independientes, los niveles de variación nucleotídica y la recombinación no deberían estar correlacionados, debido a que el entorno recombinacional no debería afectar ni al número ni a la frecuencia de las mutaciones neutras (Hudson 1983). Esto se cumpliría a excepción de que la recombinación por sí misma sea mutagénica y genere mayor nivel de polimorfismo donde mayor sea la tasa de recombinación (Yeadon y Catcheside 1999; Lercher y Hurst 2002). Bajo esta visión mutagénica de la recombinación, la divergencia ( $d$ ) debería también estar correlacionada con la recombinación (Begun y Aquadro 1992; Nachman 2001). Aunque esta correlación se ha observado en algunas especies como humanos (Filatov y Gerrard 2003; Hellmann *et al.* 2003; Spencer *et al.* 2006), ratones (Huang *et al.* 2005) o *C. elegans* (Cutter y Payseur 2003), no se ha observado en *Drosophila* para un número elevado de genes en estudios diferentes (Begun y Aquadro 1992; Presgraves 2005; Casillas *et al.* 2007; Begun *et al.* 2007; Shapiro *et al.* 2007) incluso en aquellos en los que las medidas de divergencia y entrecruzamiento eran de gran precisión y resolución (McGaugh *et al.* 2012) (Tabla 1.8).

Bajo la hipótesis de sitios ligados, tanto los arrastres por selección positiva (Gillespie 2001) como los arrastres por selección negativa (Charlesworth *et al.* 1993) podrían generar esta correlación debido a que los genes situados en regiones de mayor tasa de recombinación serían más propensos a escapar del efecto de la selección sobre los sitios selectivos adyacentes (Begun y Aquadro 1992). Además, bajo estos modelos, no se espera una correlación entre la divergencia y la recombinación. Se ha demostrado teóricamente que la selección ligada no afecta los niveles de fijación neutra a largo plazo (Birky y Walsh 1988). La interferencia selectiva de *loci* ligados se puede considerar a efectos teóricos como una reducción en el censo efectivo de la población de dicha región, y como se ha comentado anteriormente (sección 1.1.1), la tasa de evolución neutra es independiente del censo efectivo de la población. Es decir, que tras un arrastre por selección positiva, el polimorfismo se reducirá proporcionalmente a la fuerza de la selección y la tasa de entrecruzamiento pero los niveles de divergencia no se verán afectados (Birky y Walsh 1988). Por lo tanto, la hipótesis más plausible a la correlación positiva recurrentemente observada en *Drosophila* (Tabla 1.8) entre los niveles de polimorfismo y la tasa de recombinación es la acción repetida de la selección ligada en el genoma.

**Tabla 1.8.** Relación entre la recombinación y las medidas de diversidad y divergencia

Especie	Taxón externo	Tiempo divergencia (My)	Correlación, <i>c</i> y <i>p</i>	Correlación, <i>c</i> y <i>k</i>	Referencias
<i>Humanos</i>	<i>Chimpancés</i>	5-6	Positiva P = 0,060 r = 0,678	Positiva P = 0,006 r = 0,866	Bussell <i>et al.</i> 2006
<i>Humanos</i>	<i>Chimpancés</i>	5-6	Positiva P < 0,001 r = 0,735	No correlación P = 0,570 r = 0,141	Nachman 2001
<i>Mus musculus domesticus</i>	<i>Mus caroli</i>	2,5	Positiva P = 0,097 r = 0,902	No correlación P = 0,566 r = 0,434	Nachman 1997
<i>Drosophila melanogaster</i>	<i>Drosophila simulans</i>	2,5-3,4	Positiva P = 0,0007 r = NA	No correlación P = 0,983 r = NA	Begun y Aquadro 1992
<i>Drosophila simulans</i>	<i>Drosophila melanogaster</i>	2,5-3,4	Positiva P = $8,5 \cdot 10^{-5}$ r = NA	No correlación P = 0,030 r = NA	Begun <i>et al.</i> 2007
<i>Drosophila pseudoobscura</i>	<i>Drosophila persimilis</i>	0,5-1	Positiva P < 0,016 r ~0,500	Positiva P = 0,0006 r = 0,635	Kulanthinal <i>et al.</i> 2008
<i>Drosophila pseudoobscura</i>	<i>Drosophila miranda</i>	2-3	Positiva o no correlación Intrón: P = 0,004 r = 0,333 Intergénico: P = 0,258 r = 0,133	No correlación P > 0,128	Stevison y Noor 2010
<i>Zea mays mays</i>	<i>Zea mays parviglunis</i>	0,0075-0,009	Positiva P = 0,007 r = 0,650	No correlación P = NA r = NA	Tenaillon <i>et al.</i> (2001, 2002, 2004)

Correlación entre la tasa de recombinación (*c*) y el polimorfismo (*p*) o divergencia (*k*) para diversas especies. Cuando la correlación entre polimorfismo y recombinación es positiva pero la correlación entre divergencia y recombinación no lo es, se considera que hay evidencias que la selección natural afecta a los patrones de polimorfismo. Si, por el contrario la correlación es positiva para ambas comparaciones, la teoría neutra podría explicar los patrones de variación observados suponiendo un papel mutagénico de la recombinación. My: tiempo de divergencia en millones de años, r: recombinación, p: polimorfismo, k: divergencia. P: valor probabilidad (significación) de la prueba, r: valor de la correlación lineal. NA significa que los datos no se mostraron o no estaban disponibles en los trabajos originales. Resumen de la tabla de Smukowski y Noor 2011.

---

Aguadé *et al.* (1989) observaron por primera vez que genes situados en regiones donde la recombinación era baja mostraban niveles de heterocigosidad reducida. Posteriormente, Begun y Aquadro (1992) mostraron la existencia de una correlación positiva entre el polimorfismo y la tasa de recombinación. Esta correlación se ha descrito en muchos organismos además de en *Drosophila*, como en humanos, ratones, nematodos, maíz o tomate (Tabla 1.8; Smukowski y Noor 2011) hasta el punto que algunos autores consideran esta correlación como uno de los patrones más universales de la genética de poblaciones (Hahn 2008).

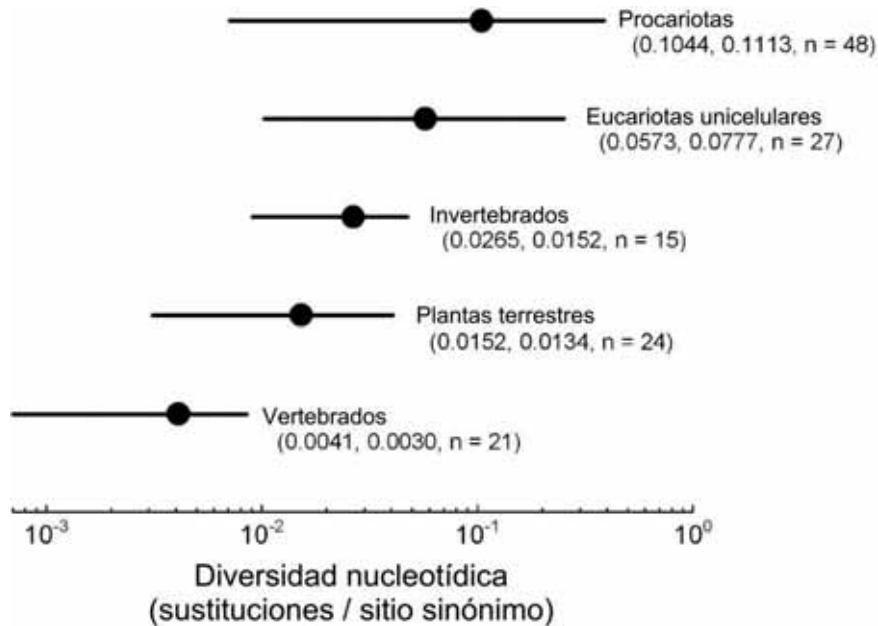
Los resultados, al menos en *Drosophila*, parecen indicar que los efectos genómicos de la selección ligada son tan significativos que se debería proponer un modelo en el que los arrastres selectivos son tan comunes que dominaran la dinámica de los sitios neutros en detrimento de la deriva genética (Gillespie 2000). Este modelo se denomina *draft* genético y fue originalmente propuesto por Gillespie (Cuadro 4). En general, como ya se ha mencionado, la selección ligada causa una reducción en  $N_e$  que reduce los niveles de polimorfismo neutro y una reducción de la eficiencia de la selección tanto de la selección negativa como de la selección adaptativa. Sin embargo, cuestiones como qué valores de recombinación se requieren para liberar al genoma de la interferencia selectiva o si la selección ligada afecta a todo el genoma por igual restan todavía sin responder.

#### 1.4.2. La selección ligada y la paradoja de la variación

Bajo la teoría neutralista, las medidas de polimorfismo genético están correlacionadas linealmente con el censo efectivo de la población (sección 1.1.1). Las primeras estimas de variación genética de la era electroforética mostraban que la diversidad está débilmente correlacionada con el *censo* de la población. A esta observación se le denominó la *paradoja de la variación* (Lewontin 1974). En un primer estudio, los niveles intraespecíficos de la variación alozímica parecían casi independientes del censo de efectivo de las especies (Lewontin 1974). Posteriormente, se describió una correlación débil entre ambas medidas (Nei 1983), aunque en las poblaciones grandes los valores de variación eran claramente inferiores a lo esperado bajo la teoría neutralista. Medidas de variación nucleotídica de cientos de especies filogenéticamente diversas, desde las procariotas hasta los vertebrados, muestran que a pesar de que el *censo* de población varía en varios órdenes de magnitud, el intervalo de diferencias en diversidad nucleotídica sólo abarcan dos órdenes de magnitud (Lynch 2006, Figura 1.12). Entre los eucariotas, no hay correlación entre los niveles de variación en el DNA mitocondrial y el *censo* de población (Bazin *et al.* 2006) y tan sólo hay una correlación débil entre los genes nucleares y el *censo* de población (Lynch 2006), a pesar de que la teoría neutralista predice una relación lineal. Sin embargo, cuando sólo se centran en especies con un *censo* de población supuestamente pequeño, como son los mamíferos placentarios (euterios), se observa una correlación positiva entre la variación nucleotídica mitocondrial y el valor del censo efectivo



(Mulligan *et al.* 2006). Por tanto, estos resultados estarían de acuerdo con el modelo de *draft* genético de Gillespie: en aquellas especies que presentan un censo efectivo reducido la variación genética estará gobernada por el equilibrio mutación-deriva. Pero una vez superado un umbral de censo poblacional, especies con censos efectivos muy diversos no variarán sustancialmente en sus niveles de diversidad genética debido a procesos repetidos de arrastre por selección natural (Cuadro 4, Figura 1.13).

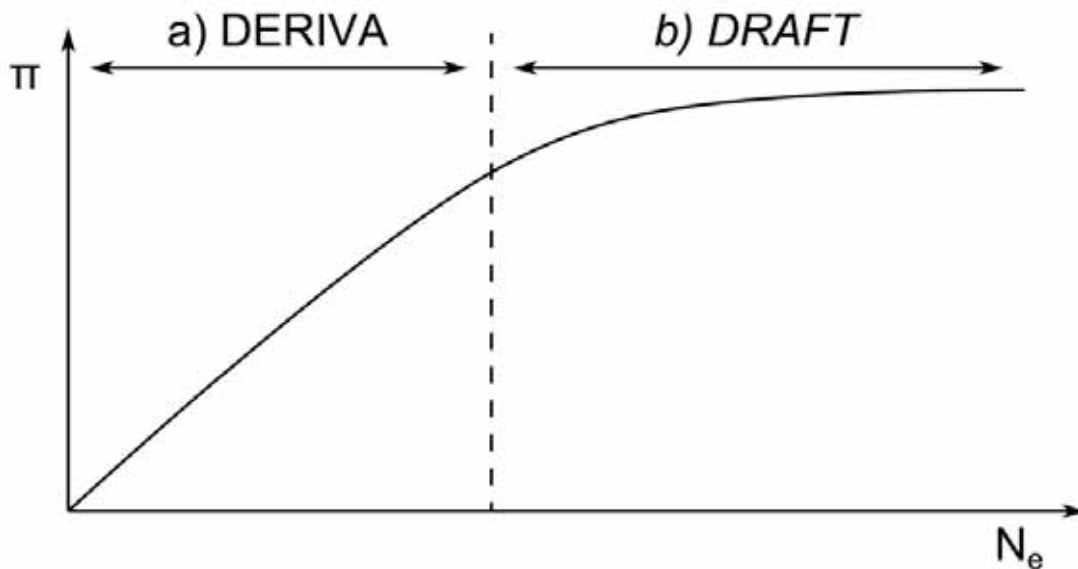


**Figura 1.12.** *La paradoja de la variación genética*

Promedio de la diversidad nucleotídica (medido como el número de sustituciones por sitio) en los sitios sinónimos de los genes codificadores de una serie de especies filogenéticamente diversas (incluyendo especies procariotas y eucariotas). Entre paréntesis se muestran la media, desviación estándar y el número de géneros que conforman los distintos grupos taxonómicos. Las barras horizontales son las estimas de los intervalos de confianza del 95%. Figura tomada de Lynch 2006.

Si suponemos que los sitios polimórficos evolucionan independientes entre sí, se debería observar una correlación positiva entre la diversidad nucleotídica y el censo efectivo de la población. Sin embargo, si se considera que los sitios están ligados, los modelos teóricos predicen que la diversidad (polimorfismo) nucleotídica y el valor de censo efectivo estarán desacoplados (*draft* genético de Gillespie, 2001) o bien que la relación entre estas dos variables será muy débil (modelo de arrastre por selección negativa de Charlesworth *et al.* 1993). Es decir, que especies con un censo efectivo menor pueden tener una intensidad de selección tan baja como para que la unidad de selección sea a la práctica cada sitio polimórfico y actúen por tanto como si fueran independientes, haciendo que los posibles arrastre por selección tengan un efecto menor tal y como predice la teoría neutralista (Figura 1.13a). Sin embargo, en especies donde el censo efectivo es mayor, la intensidad y frecuencia de la selección puede ser suficientemente importante como para que la unidad de selección sea el conjunto de sitios que

cosegregan y los arrastres por selección natural gobiernen la variación genética observada (Figura 1.13b).



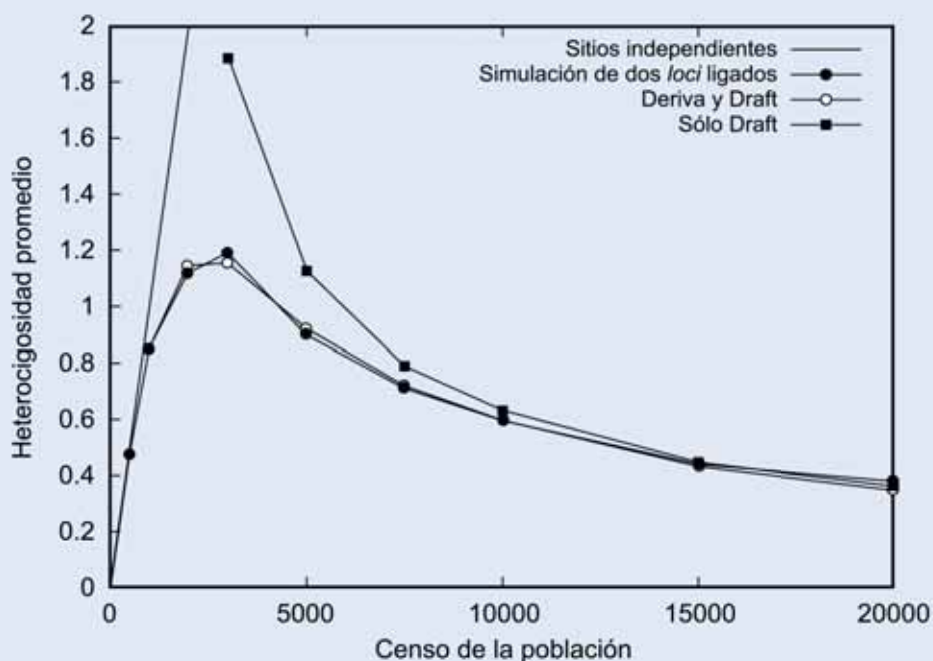
**Figura 1.13.** Esquema de la relación entre la diversidad nucleotídica y el censo efectivo de la población

(a) Cuando el censo efectivo de la población ( $N_e$ ) es bajo, la deriva genética es la fuerza dominante y la diversidad ( $\pi$ ) incrementa linealmente con  $N_e$ . (b) Sin embargo, para valores de  $N_e$  suficientemente altos es el *draft* genético el factor determinante y  $\pi$  no se correlaciona con  $N_e$ .

Recientemente se ha observado que la variable que mejor predice la diversidad nucleotídica entre diversos linajes de animales (76 especies) es la inversión en el cuidado parental de las especies. Las especies con mayor cuidado parental (tortugas, mamíferos, pájaros, caballitos de mar, hormigas o abejas entre otros) muestran menor diversidad nucleotídica mientras que las especies con una inversión mínima en sus descendientes (mejillones, ascidias y gorgonias entre otros) tienen unos niveles de diversidad mucho mayores (Figura 1.15, Romiguier *et al.* 2014). Es decir, que gracias al análisis de nuevos datos genómicos analizados conjuntamente con variables ecológicas precisas de diversas especies no modelo, Romiguier y colaboradores (2014) ponen de manifiesto que la inversión en el cuidado parental, variable relacionada con el censo efectivo de las especies, explica mejor la relación con la diversidad nucleotídica que las variables que afectan a la historia de las especies, ya sea las fluctuaciones demográficas o la acción variable de la selección natural.

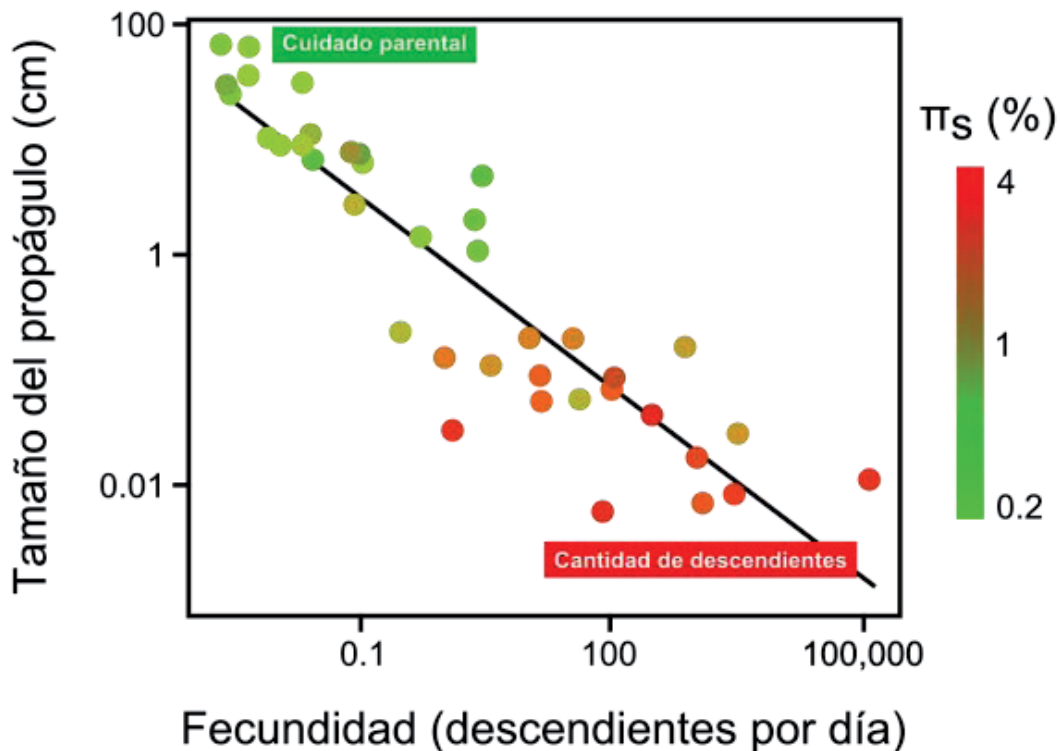
#### Cuadro 4. *Draft* genético

John Gillespie revisó la hipótesis del arrastre por selección positiva y desarrolló un modelo estocástico que denominó *draft* genético (Gillespie 2000<sup>a</sup>; Gillespie 2000b y Gillespie 2001). El *draft* genético da lugar a una dinámica aparentemente similar a la deriva genética, pues una de sus consecuencias es la pérdida de variación genética en las poblaciones (Figura 1.14). En contraposición a lo que se espera bajo la deriva genética, la reducción de la variación incrementa con el *censo* de la población bajo el modelo del *draft genético* (Figura 1.14). Esto es debido a que la tasa de arrastres por selección positiva incrementa con el *censo* de la población a consecuencia de que cuanto mayor sea la población (i) mayor será la probabilidad de sufrir mutaciones adaptativas y (ii) la selección será más eficiente permitiendo incluso la fijación de alelos débilmente adaptativos. Por tanto, con el incremento del *censo* efectivo, la diversidad nucleotídica aumentará pero también lo hará la tasa de los arrastres por selección positiva reduciendo la diversidad. Así, cuando el *censo* efectivo sea suficientemente grande, el *draft* genético dominará la dinámica de la variación y la diversidad nucleotídica será independiente del *censo* efectivo de las poblaciones (Figura 1.14). Mediante este elegante modelo Gillespie trata de explicar el desacoplamiento del *censo* efectivo de los niveles de diversidad genética de las poblaciones (Gillespie 2004; Lynch 2007).



**Figura 1.14.** Simulación del efecto de la selección ligada en la heterocigosidad de un locus

Heterocigosidad promedio de un sitio neutro ligado a un locus fuertemente seleccionado. Los valores de la línea marcada como “Sólo Draft” se han obtenido de la ecuación  $2\mu/\rho_N E\{y^2\}$ , siendo  $\rho_N$  la tasa de sustitución en el locus fuertemente seleccionado e  $y$  la frecuencia final del alelo neutro ligado al locus fuertemente seleccionado. Esta ecuación es el equivalente a  $4N\mu$  (modelo basado en deriva) bajo un modelo de *draft* genético. Los valores de la línea marcada como “Deriva y Draft” se han calculado de la ecuación  $4N\mu/1+2N\rho_N E\{y^2\}$  donde se combinan los modelos de deriva y *draft* genético. La línea “Sitios independiente” hace referencia a la simulación de un locus neutro no ligado o aislado. Figura adaptada de Gillespie (2001).



**Figura 1.15.** Correlación de las características del ciclo vital y la diversidad genética de distintas especies

Efecto de la fecundidad por día (eje x) y el tamaño del propágulo (eje y) sobre la diversidad nucleotídica (escala de colores). Cada punto es el promedio de una familia taxonómica.  $r^2 = 0,69$ , valor- $p < 10^{-6}$ . El tamaño del propágulo es el tamaño que tiene el individuo cuando abandona a sus progenitores y se dispersan, ya sea huevo o juvenil, dependiendo de la especie. Figura tomada de Romiguier *et al.* 2014.

### 1.5. *Drosophila melanogaster* como organismo modelo

La mosca de la fruta *Drosophila melanogaster* es uno de los modelos experimentales más exitosos de la investigación genética (Roberts, 2006). Se introdujo a principios del siglo XX como herramienta para el análisis genético y resultó ser crucial en los primeros pasos de la genética (Morgan 1915; Muller 1927) y también para la rama de la genética de poblaciones. Desde entonces se ha mantenido a la vanguardia de la investigación, proporcionando información empírica sobre los factores que afectan la variación genética en las poblaciones naturales (Ayala *et al.* 1974; Singh y Rhomberg 1987; Powell 1997), la construcción del plan corporal de los animales (Lewis 1978; Nüsslein-Volhard y Wieschaus 1980) y la función del sistema nervioso (Ivanov *et al.* 2004), entre otros. La gran cantidad de información funcional disponible y las múltiples técnicas de manipulación genética que existen, hace que *D. melanogaster* sea un organismo muy adecuado para el estudio de las fuerzas evolutivas que determinan la variación genética de las poblaciones naturales. Podemos afirmar por lo tanto que *Drosophila* ha sido y continúa siendo el organismo modelo por excelencia de la genética de poblaciones.

---

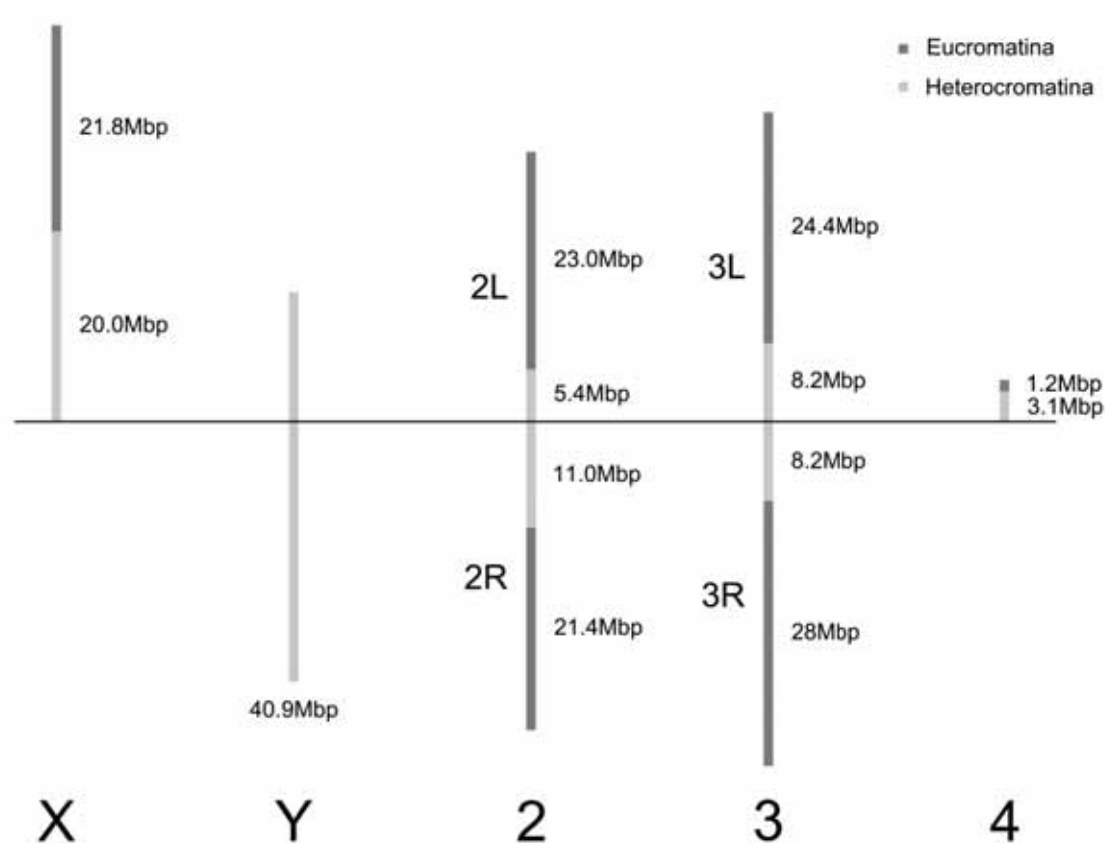
### 1.5.1. Características del genoma

Tras la secuenciación del genoma de la levadura *Saccharomyces cerevisiae* (Goffeau *et al.* 1995) y del nematodo *Caenorhabditis elegans* (*C. elegans Consortium* 1998) se secuenció por primera vez el genoma completo de *D. melanogaster*. Fue seleccionado entre los primeros organismos eucariotas a secuenciarse por completo debido a su importancia histórica en la investigación genética junto con su modesto genoma (Rubin 1996; Adams *et al.* 2000). Desde entonces, se han generado una multitud de nuevas herramientas genómicas, incluyendo bases de datos especializadas y herramientas de análisis de secuencias (Matthews *et al.* 2005; Fox *et al.* 2006; Galperin 2007) que han permitido mejorar la secuencia genómica inicial en calidad y riqueza (Ashburner y Bergman 2005), en la determinación de regiones ambiguas como los huecos o *gaps* (Celniker *et al.* 2002; Hoskins *et al.* 2002) así como en una definición extensiva de las anotaciones funcionales (Kopczynski *et al.* 1998; Berger *et al.* 2001; Berman *et al.* 2002; Kaminker *et al.* 2002, Misra *et al.* 2002; Ohler *et al.* 2002; Carvalho *et al.* 2003; Bergman *et al.* 2005; Tupy *et al.* 2005; Hoskins *et al.* 2011; Hoskins *et al.* 2015). *D. melanogaster* sigue siendo hoy día uno de los genomas eucariotas mejor anotados.

El genoma de *Drosophila melanogaster* mide aproximadamente 180Mb repartidos en 5 cromosomas; X, Y, 2, 3 y el 4 o *dot* (Figura 1.16) (Hoskins *et al.* 2002). Se han anotado 17.716 genes (versión 6.05 del genoma de referencia). Se estima que el genoma euromático de *Drosophila melanogaster* consta de 120Mb, de los cuales más del 75% está formado por regiones intrónicas e intergénicas, es decir, por regiones no codificadoras (Misra *et al.* 2002).

### 1.5.2. Filogenia

Los artrópodos son uno de los taxones animales más diversos y exitosos, pues representan aproximadamente el 75% de todas las especies animales. En el linaje de los insectos se encuentra el taxón *Drosophila* que lo conforman más de 2.000 especies (Powell 1997). Estudios filogenéticos basados en la tasa de mutación genómica han determinado que los principales linajes dentro del género *Drosophila* divergieron hace unos 40 - 62 MYA (Russo *et al.* 1995; Tamura *et al.* 2004) (Figura 1.17). Uno de los linajes lleva al subgénero *Sophophora*, formado por ~330 especies conocidas, entre las que se encuentra la especie *D. melanogaster* junto con especies como *D. simulans* o *D. yakuba* utilizadas habitualmente como taxones externos del organismo modelo. El otro linaje dio lugar a los subgéneros *Drosophila* e *Idiomya* (*Drosophilas* Hawaianas), los cuales están formados por ~1.100 y ~380 especies identificadas, respectivamente.

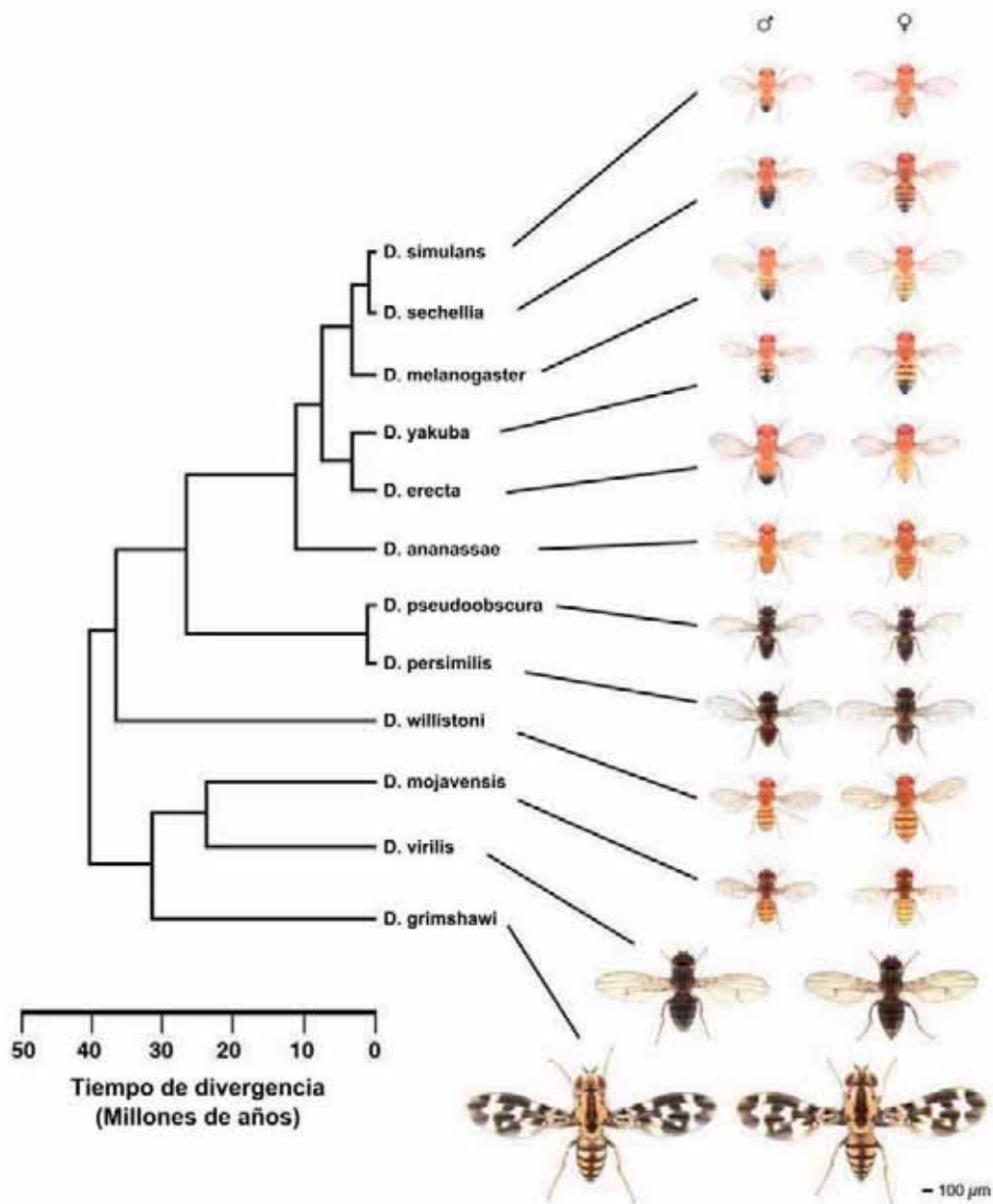


**Figura 1.16.** Esquema de los cromosomas de *D. melanogaster*

Se representan las regiones euromáticas (gris oscuro) y heterocromáticas (gris claro) de los distintos cromosomas.

### 1.5.3. Demografía

*D. melanogaster* es una especie originaria del África subsahariana que se ha expandido recientemente al resto del mundo convirtiéndose así en una especie cosmopolita (Lachaise *et al.* 1988; David y Capy 1988; Begun y Aquadro 1993; Andolfatto 2001; Stephan y Li 2007). La expansión desde África hacia Europa se ha datado aproximadamente entre hace 10.000 y 19.000 años, lo que equivale a 0,1 - 0,4  $N_e$  generaciones (Li y Stephan 2006; Thornton y Andolfatto 2006; Duchon *et al.* 2013). Sin embargo, las primeras evidencias de que *D. melanogaster* llegó a América del Norte son de hace menos de 200 años (Johnson 1913; Sturtevant 1920; Keller 2007). Debido a que en aquella época la fauna díptera estaba muy bien descrita, es improbable que los entomólogos pasaran por alto a *D. melanogaster* durante muchos años (Keller 2007). En menos de 25 años, esta especie se propagó por todo el continente convirtiéndose así en uno de los dípteros más comunes en América del Norte (Howard 1900).



**Figura 1.17.** *Filogenia del género Drosophila*

Árbol filogenético de las 12 especies del género *Drosophila* cuyos genomas fueron los primeros secuenciados y analizados (Clark et al 2007, *Assembly/Alignment/Annotation of 12 related Drosophila species*; <http://rana.lbl.gov/drosophila/>).

Inicialmente se demostró que la variación encontrada en cualquier población no africana era menor a la encontrada en África (Begun y Aquadro 1993; Andolfatto 2001) y parecía por tanto que *D. melanogaster* había sufrido una reducción en su *censo* de población (Akashi 1996). Esto sugería que la propagación fuera de África de las distintas poblaciones fue precedida por un cuello de botella (Begun y Aquadro 1993; Andolfatto 2001; Li y Stephan 2006; Thornton y Andolfatto 2006). Además, analizando la diferenciación de las poblaciones se puso de manifiesto que las poblaciones americanas eran genéticamente más cercanas a las poblaciones

---

africanas que a las europeas (Caracristi y Schlötterer 2003; Baudry *et al.* 2004; Haddrill *et al.* 2005b). Un estudio reciente sugiere que la población norteamericana es una mezcla migratoria entre poblaciones y estiman que la proporción ancestral africana es del 15% y la europea del 85% (Duchen *et al.* 2013). Con esta distribución cosmopolita se espera que las diferentes poblaciones hayan evolucionado y adaptado de forma diferente a los distintos ambientes, haciendo de *Drosophila* un organismo modelo apropiado para el estudio de la adaptación y/o de la demografía poblacional; en definitiva para la genética de poblaciones.

## 1.6. Proyecto DGRP

Hasta ahora la mayoría de los estudios de genética de poblaciones se han basado en muestras genómicas fragmentadas y no aleatorias del genoma, generando una visión parcial y muchas veces sesgada de los procesos de genética de poblaciones. El *data desideratum* para un estudio completo y no sesgado de variación genética sería disponer de una gran muestra de genomas completos de alta calidad de una misma población natural de un organismo modelo (del que existe un extenso conocimiento biológico) y la secuencia completa de un taxón externo. Esto permitiría una descripción completa y detallada de los patrones de variación a gran escala.

Durante los últimos años, a medida que el precio de los proyectos de secuenciación ha ido disminuyendo, se han comenzado a secuenciar, o resecuenciar, individuos de organismos modelo de los que previamente sólo existía un genoma de referencia secuenciado. Algunos ejemplos de este tipo de proyectos se recogen en la Para el organismo modelo *Drosophila melanogaster* se ha llevado a cabo una iniciativa internacional en la que se han obtenido las secuencias genómicas de alta calidad de decenas de individuos muestreados en una misma población (Mackay *et al.* 2012). El proyecto, denominado DGRP (*Drosophila Genetic Reference Panel*), es un recurso formidable para la comunidad drosofilista que abre una nueva dimensión al estudio de la variación genotípica y fenotípica de este organismo modelo. Para llevar a cabo este proyecto se muestrearon moscas en las cercanías de un mercado horticultor de Raleigh, Carolina del Norte (EEUU). Tras 20 generaciones de cruces hermano-hermana se obtuvieron una colección de isolíneas prácticamente consanguíneas.

Tabla 1.9. Sin embargo, en muchos de estos casos, sólo se han podido secuenciar regiones concretas del genoma o secuencias genómicas de baja calidad, secuencias borrador, o un número reducido de individuos.

Para el organismo modelo *Drosophila melanogaster* se ha llevado a cabo una iniciativa internacional en la que se han obtenido las secuencias genómicas de alta calidad de decenas de individuos muestreados en una misma población (Mackay *et al.* 2012). El proyecto, denominado DGRP (*Drosophila Genetic Reference Panel*), es un recurso formidable para la comunidad drosofilista que abre una nueva dimensión al estudio de la variación genotípica y



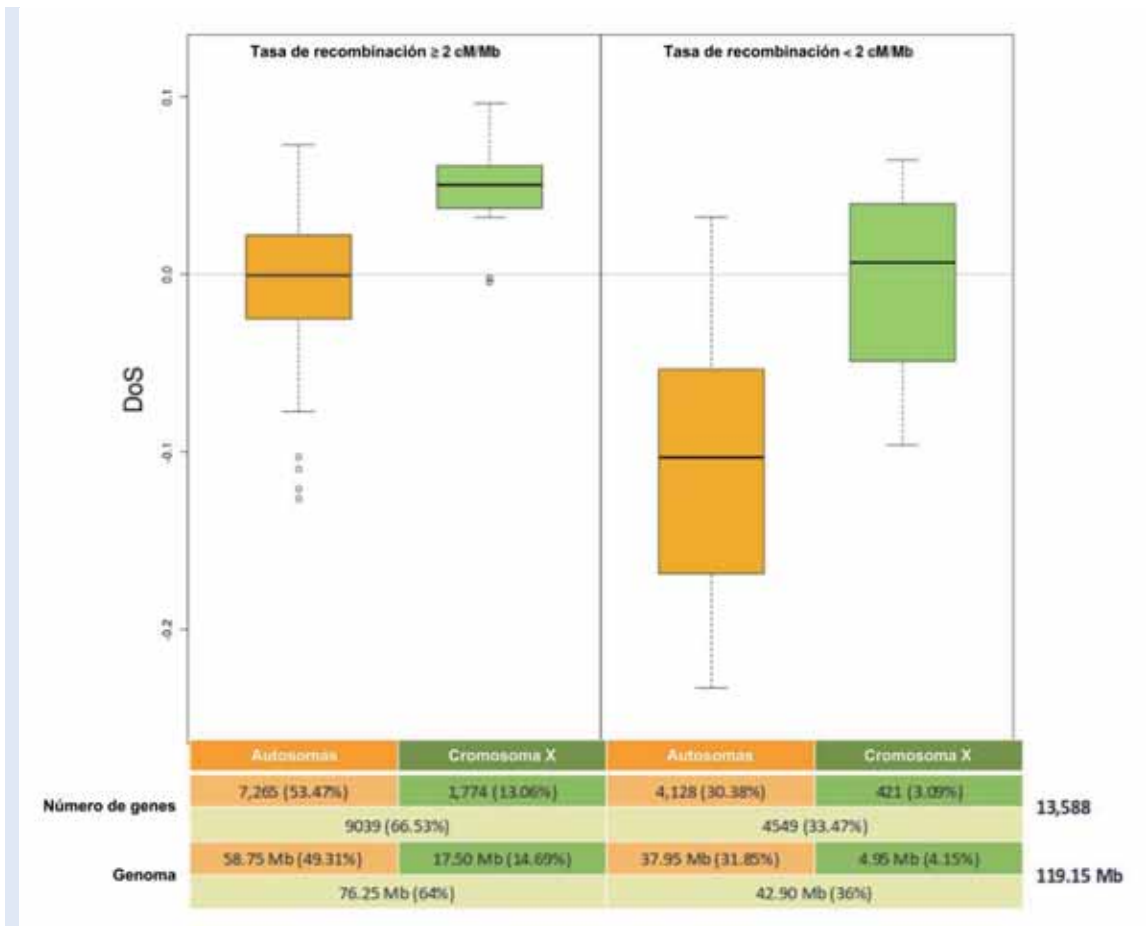
fenotípica de este organismo modelo. Para llevar a cabo este proyecto se muestrearon moscas en las cercanías de un mercado horticultor de Raleigh, Carolina del Norte (EEUU). Tras 20 generaciones de cruces hermano-hermana se obtuvieron una colección de isolíneas prácticamente consanguíneas.

**Tabla 1.9.** *Proyectos de resecuenciación en distintos organismos modelo*

Organismo modelo	Proyectos	Individuos	Cobertura	Referencias
<i>Homo sapiens</i>	1000 Genomes Project	1092	~4X	<a href="http://www.1000genomes.org">http://www.1000genomes.org</a>
<i>Mus musculus</i>	Mouse Genomes Project	17	~25X	<a href="http://www.sanger.ac.uk/resources/mouse/genomes/">http://www.sanger.ac.uk/resources/mouse/genomes/</a>
<i>Drosophila melanogaster</i>	DGRP	244	~24X	<a href="http://dgrp.gnets.ncsu.edu">http://dgrp.gnets.ncsu.edu</a>
	DPGP2	623	~25X	<a href="http://www.dpgp.org/dpgp2/DPGP2.html">www.dpgp.org/dpgp2/DPGP2.html</a>
<i>Saccharomyces cerevisiae</i>	SGRP	37	1-3X	<a href="https://www.sanger.ac.uk/research/projects/genomeinformatics/sgrp.html">https://www.sanger.ac.uk/research/projects/genomeinformatics/sgrp.html</a>
<i>Arabidopsis thaliana</i>	1001 Genomes Project	1049	~17X	<a href="http://1001genomes.org">http://1001genomes.org</a>

### Cuadro 5. Genómica de poblaciones de las secuencias DGRP

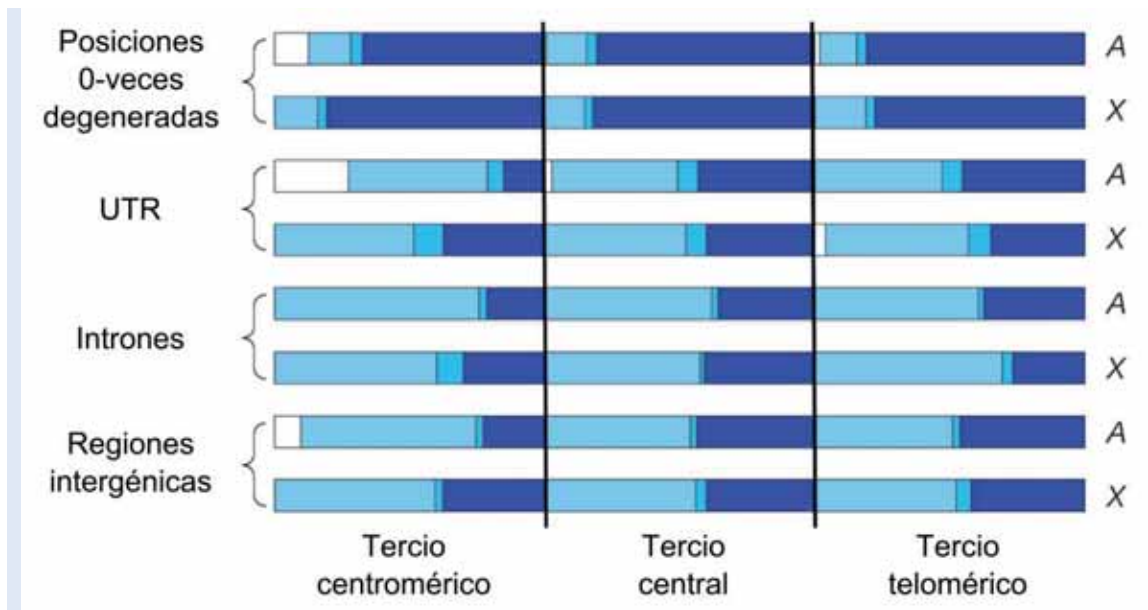
Mackay y colaboradores (2012) llevaron a cabo el estudio de genómica de poblaciones de las isolíneas de DGRP. Un resultado muy significativo surgió del análisis de la relación entre el polimorfismo y la recombinación. Los modelos evolutivos de los arrastres por selección positiva y negativa (Begun y Aquadro 1992; Charlesworth *et al.* 1993) predicen una correlación positiva entre el polimorfismo y la tasa de recombinación. La existencia de correlación se ha corroborado en este estudio, pero no se observa a lo largo de todo el genoma, sino en aquellas regiones en las que la recombinación local es inferior a 2 cM/Mb (R Spearman = 0,471, valor-p = 0). Por encima de este valor de recombinación, el polimorfismo y la tasa de recombinación son estadísticamente independientes (R Spearman = -0,004, valor-p = 0,987). La falta de correlación en las regiones por encima de 2 cM/Mb no se puede interpretar como la ausencia de selección direccional en esas regiones puesto que la selección positiva es más común en las regiones de mayor recombinación comparadas con las regiones < 2 cM/Mb de recombinación (Figura 1.18). En ningún caso se observó una correlación positiva entre la recombinación y la divergencia ( $r^2 = -0,2755$ , valor  $p = 0$ ).



**Figura 1.18.** Dirección de la selección bajo dos contextos recombinacionales ( $\geq 2$  cM/MB y  $< 2$  cM/MB)

La dirección de la selección, DoS, es un índice propuesto por Stoletzki y Eyre-Walker (2011) para cuantificar la selección: DoS positivo indica selección adaptativa, cero neutralidad y negativo alelos ligeramente deletéreos segregando. DoS entre autosomas (naranja) y el cromosoma X (verde) para regiones genómicas con una tasa de recombinación  $< 2$  cM/MB (derecha) y para regiones genómicas con una tasa de recombinación  $> 2$  cM/MB (izquierda). DoS se estimó para cada gen de forma independiente y se promedió en ventanas de 50 genes a lo largo de los brazos cromosómicos.

Además, se escudriñó el genoma en busca de evidencias de selección natural. Para ello se incorporó información de la distribución del espectro de frecuencias en el marco del MKT (sección iii), Tabla 1.3). Se estimó que un 58,5% de los sitios del genoma son neutros o casi neutros mientras que un 1,9% son ligeramente deletéreos y un 39,6% fuertemente deletéreos. Pero estas proporciones varían entre los autosomas y el X, entre clases de sitios y en función del tercio cromosómico en el que se encuentren (Figura 1.19). En el tercio centromérico de los autosomas la proporción de sitios fuertemente deletéreos está reducido y la fracción de sitios neutros aumentado. Esto no ocurre en el cromosoma X, en el que no hay diferencias entre las regiones genómicas. En cuanto a la selección positiva, en promedio el 25,2% de los sitios fijados entre *D. melanogaster* y *D. yakuba* son adaptativos. La mayor parte de las sustituciones adaptativas se dan en el cromosoma X (36%) seguidas de las regiones no centromérica de los autosomas (25%). En las regiones centroméricas de los autosomas prácticamente no hay evidencias de selección adaptativa (1%).



**Figura 1.19.** Regímenes de selección en el genoma de *D. melanogaster*

La fracción de nuevas mutaciones que están sujetas a diferentes regímenes de selección por clase de sitio y región cromosómica para los autosomas (A) y el cromosoma X (X). Regímenes de selección: sitios muy deletéreos (azul oscuro), sitios ligeramente deletéreos (azul), sitios neutros antiguos (azul claro) y sitios neutros recientes (blanco) (Tabla 1.3, sección iii).

Para secuenciar los genomas se han combinado dos métodos de secuenciación masiva con características complementarias. Por un lado, la tecnología Illumina con la que se obtienen secuencias cortas (< 100pb) pero que son particularmente adecuadas para detectar SNPs, y por otro secuencias obtenidas con la tecnología Roche 454, más largas que permiten detectar de forma más precisa cambios estructurales. En la primera fase de este proyecto (Freeze 1; Mackay *et al.* 2012) se secuenciaron 158 líneas de Illumina y 39 de 454, y estas secuencias son las que estaban disponibles a lo largo de la realización de esta tesis. La secuencia de referencia de *Drosophila* se utilizó como patrón para cartografiar las lecturas. Por lo tanto, sólo se pudieron analizar las variaciones de un único nucleótido o SNPs. Recientemente, se ha publicado la segunda fase del proyecto (Freeze 2; Huang *et al.* 2014) en el que se amplió el número de líneas secuenciadas a 454 y se han analizado variantes estructurales como las inserciones y deleciones a lo largo de todo el genoma. La llegada de estos nuevos datos ha sido posterior a la elaboración de este trabajo doctoral y por tanto no se han analizado en esta tesis.

El objetivo principal de proyecto DGRP es entender la base genética que subyace a la variación fenotípica de las poblaciones. Es decir, entender la relación entre el genotipo y el fenotipo de los caracteres complejos (Mackay *et al.* 2012). El DGRP es un recurso para la comunidad de drosofilistas con tres objetivos principales: (i) llevar a cabo estudios de asociación de los QTLs y posicionarlos en regiones específicas del genoma, (ii) crear una

---

colección de SNPs y variantes estructurales de *Drosophila melanogaster*, (iii) crear un banco de pruebas en los que contrastar los métodos estadísticos usados en los estudios de asociación para caracteres que afectan a enfermedades humanas. Entre otros caracteres cuantitativos, se estudiaron los fenotipos: resistencia al estrés por inanición (*starvation stress*), tiempo de recuperación al coma por frío (*chill coma recovery time*), y respuesta a estímulos repentinos (*startle response*), para intentar determinar las variantes genéticas asociadas a estos fenotipos. Para ello se realizó un estudio de asociación a lo largo del genoma (GWA; *Genome-wide association study*). El GWA de los tres caracteres cuantitativos reveló cientos de nuevos genes candidatos, lo que subraya nuestra ignorancia sobre las bases genéticas de los caracteres complejos. La mayoría de las variantes asociadas a los fenotipos se hallaban a baja frecuencia en la población. De hecho se observó una correlación negativa entre la frecuencia de una variante y su efecto. Los modelos de regresión que incorporan los SNPs significativos explican la mayoría de la variación fenotípica del carácter (entre el 72% - 85%), a diferencia de lo que ocurre en estudios de asociación en humanos donde los SNPs significativos solo explican una pequeña fracción de la variación (Manolio *et al.* 2009).

Además de su interés para estudios de asociación genotipo-fenotipo, el proyecto DGRP es un recurso único para llevar a cabo el análisis más completo de genómica de poblaciones hasta la fecha. El grupo de investigación dirigido por el Dr. Barbadilla, bajo cuya dirección se ha realizado esta tesis, se propuso llevar a cabo los siguientes objetivos: (i) crear y mantener un visualizador genómico del mapa de las secuencias polimórficas de alta resolución de *D. melanogaster* y (ii) efectuar el análisis de genómica de poblaciones de todas las secuencias genómicas disponibles. Los resultados principales de este análisis, que no forman parte de esta tesis, se recogen en el Cuadro 5.

## 1.7. Objetivos

Esta tesis es un proyecto de genómica de poblaciones que quiere probar la hipótesis que la recombinación es un factor clave para explicar los patrones de variación y evolución de los genomas. En concreto, se quiere demostrar que los valores variables que toma la recombinación a lo largo del genoma determinan la importancia relativa de la selección natural y la deriva genética como fuerzas moduladoras de la variación genética en diferentes regiones del genoma. Para este fin se llevaron a cabo los siguientes dos objetivos principales:

### 1.7.1. Descripción y explicación de los patrones de variación en los genes codificadores de *Drosophila melanogaster*

El primer gran objetivo abordado, que se llevó a cabo durante el primer periodo de la tesis, fue la descripción y explicación de los patrones de variación a escala genómica. Para ello se definieron los siguientes objetivos específicos:

- 
- i) Descripción de la variabilidad genómica (polimorfismo y divergencia) en una población norteamericana de *Drosophila melanogaster* (las secuencias genómicas del proyecto DGRP).
  - ii) Comparación de los patrones de variación de los cromosomas autosómicos frente al cromosoma X en las distintas regiones cromosómicas y para distintas clases funcionales de sitios.
  - iii) Evaluación de la importancia de los factores genómicos (mutación, recombinación, densidad génica) como moduladores de la variación nucleotídica a lo largo del genoma.

### 1.7.2. Efecto de la recombinación sobre los patrones de variación nucleotídica

En la segunda parte de esta tesis, se llevó a cabo un análisis en profundidad de la relación entre la variación nucleotídica putativamente neutra, las de las posiciones 4 veces degeneradas, y la tasa de recombinación, uno de los factores genómicos moleculares que consideramos crucial para explicar los niveles de variación neutra. Para ello se desarrollaron los siguientes objetivos específicos:

- i) Confirmar la observación de Mackay y colaboradores (2012) que existe un umbral de recombinación (alrededor de 2 cM/Mb) en el genoma de *D. melanogaster* a partir del cual la correlación entre el polimorfismo sinónimo y la recombinación desaparece y proponer una explicación evolutiva para la existencia de ese umbral.
- ii) Búsqueda y estimación de un valor umbral de recombinación a partir del cual la recombinación y el polimorfismo neutro se desacoplan. Este estudio se realizó para cada brazo cromosómico independientemente.
- iii) Elaboración de un mapa genómico de bloques de selección ligada.
- iv) Validación del mapa de bloques de selección ligada mediante (a) la descripción de las propiedades de la dinámica evolutiva de los bloques (LD,  $r/\mu$ ), (b) la comprobación de las predicciones teóricas en función si hay o no selección ligada, y (c) el análisis comparativo de los diferentes mapas de bloques de selección ligada que resultan de los distintos tamaños (kb) de las ventanas de observación.



---

## Parte 2

# Materiales y Métodos





---

## 2. Materiales y métodos

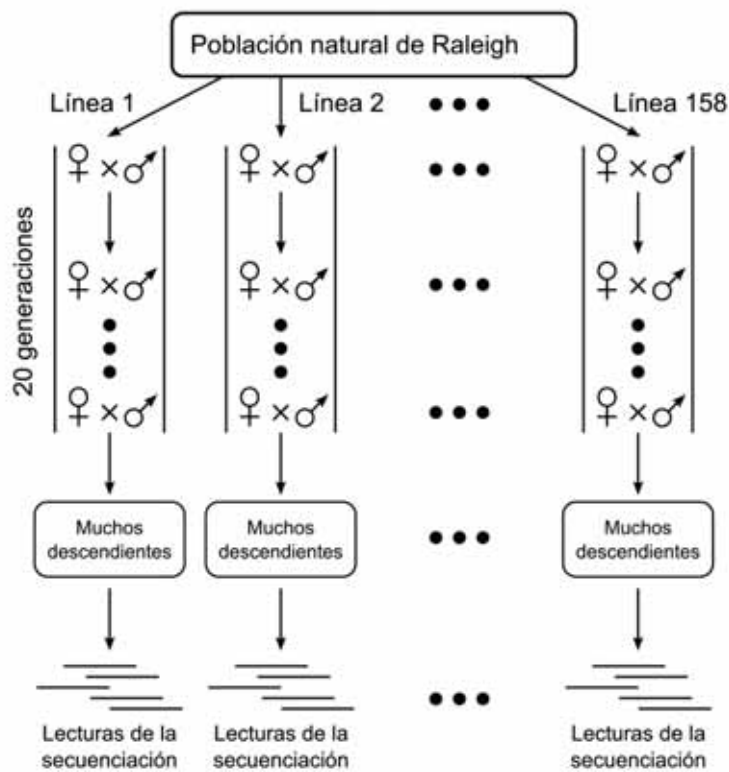
### 2.1. Secuencias genómicas

Esta tesis se inició un poco antes de que estuvieran ensambladas y accesibles para su análisis las secuencias de un conjunto de 168 líneas consanguíneas individuales de *Drosophila melanogaster* (*The Drosophila Genetic Reference Panel* (DGRP) Freeze 1; Mackay *et al.* 2012). La disponibilidad de 168 genomas de una población natural de *D. melanogaster*, junto a los genomas secuenciados de las especies cercanas a *D. melanogaster*, representaban una oportunidad sin precedentes para llevar a cabo un estudio exhaustivo de genómica de poblaciones.

Las líneas se obtuvieron tras 20 generaciones de cruces endogámicos hermano-hermana a partir de una muestra de hembras grávidas recolectadas de una población natural de Raleigh, Carolina del Norte (EEUU) (Figura 2.1). Con esta aproximación se pretende que la variabilidad genética de la población natural, capturada en la muestra aleatoria de hembras, se fije al azar en el conjunto de líneas homocigóticas que se han obtenido y secuenciado. Este procedimiento de obtener líneas endogámicas tiene la gran ventaja de soslayar el problema de genotipar heterocigotos en la secuenciación de genomas de organismos diploides. El coeficiente de consanguinidad ( $F$ ) esperado después de las 20 generaciones es de 0,986 (Falconer y Mackay 1996), es decir, teóricamente casi el 99% de los heterocigotos se habrán fijado en una u otra línea, y sólo 1% se SNPs segregarán como heterocigotos residuales en cada línea.

Una limitación a esta aproximación metodológica son los alelos deletéreos que segregan a baja frecuencia en la población natural o los alelos letales recesivos que se encuentran en heterocigotos o heterocariotipos para inversiones. Todas estas variantes deletéreas o letales se verán sometidas a un proceso tendente a su eliminación o purga genética (García-Dorado 2012) en el transcurso de las 20 generaciones de cruces endogámicos. El efecto neto sobre la estimación de la variación es que las variantes bajo selección purificadora estarán subrepresentadas en las secuencias genómicas obtenidas. Además, es posible que las presiones selectivas que actúan sobre las variantes alélicas en las condiciones artificiales de obtención de las isolíneas sean distintas a las del ambiente natural. Estas limitaciones no las consideramos graves porque el procedimiento seguido de iniciar cada generación a partir de un único cruce hermano-hermana en cada línea supone un cuello de botella muy intenso, reduciendo el censo de población a su mínima expresión en una especie sexual,  $N = 2$  (*a Noah's Ark sampling*, Spiess 1989). Bajo este régimen de endogamia extrema, la deriva es la fuerza dominante en la segregación de los alelos, siendo mucho más decisiva que la selección deletérea. Sólo variantes letales o los alelos muy deletéreos serán purgados durante la obtención de las líneas endogámicas (García-Dorado 2012). Una de las expresiones más conocidas de los modelos que

analizan conjuntamente la selección y la deriva es  $2s = 1/2N_e$  (Lynch 2007), que define la condición en la que la selección y la deriva tienen la misma influencia sobre las frecuencias alélicas (donde  $s$  es el coeficiente de selección). Si consideramos en nuestro caso  $2N_e = 4$ , se deduce de la expresión que sólo para valores  $|s| \gg 1/8$ , aplicable a alelos muy deletéreos o letales, las variantes se verán afectadas por la purga genética. El resto de las variantes, que supuestamente son la gran mayoría de la variación fijada en las aislneas, constituye una muestra representativa y aleatoria de las variantes existentes en la población natural en el momento en que la muestra fue tomada, exceptuando únicamente los alelos letales o muy deletéreos.



**Figura 2.1.** Diseño experimental seguido para la obtención y secuenciación de las líneas DGRP

Cada línea DGRP fue fundada por una hembra grávida capturada en las afueras de Raleigh, Carolina del Norte. Cada nueva generación resulta del cruce de una pareja macho y hembra de la generación anterior. Cada línea DGRP se obtuvo después de 20 generaciones hermano-hermana. De cada línea se extrajo DNA de un acervo de 500-1000 moscas que fueron secuenciadas. Figura tomada de Stone (2012).

Otra fuente potencial de error del supuesto que estamos estimando la variación natural a partir de las variantes de las secuencias de las aislneas es la aparición de nuevas variantes o SNPs por mutación durante el proceso de obtención de las aislneas a lo largo de las 20 generaciones. Si la frecuencia de estas mutaciones fuera relativamente importante respecto a las variantes naturales preexistentes, obtendríamos una sobrestimación de la frecuencia de variantes únicas, los *singletons*, en la población natural. Si suponemos que  $2N = 4$ , la tasa de mutación

---

por nucleótido y generación en *D. melanogaster* es de  $10^{-9}$  (Sharp y Li 1989; McVean y Vieira 2001; Keightley *et al.* 2009; Schrider *et al.* 2013), y el genoma mide 119,03MB, se generarán aproximadamente  $\sim 0,48$  nuevas mutaciones por generación y línea. Como sabemos que se realizaron 20 rondas de apareamiento (20 generaciones), habrán aparecido  $\sim 10$  mutaciones *de novo* de las cuales  $\frac{1}{4}$  se habrán fijado en promedio al final de las 20 rondas. Luego, sólo 2,5 variantes por línea será variación no natural. Si tenemos 158 líneas en las que este proceso puede ocurrir, habrá un promedio de 395 variantes *de novo* de los aproximadamente 4.800.000 SNPs descubiertos en las secuencias. O sea, sólo uno de cada 11.392 SNP catalogados será *de novo* y alrededor de uno de cada 3.500 *singletons*. En consecuencia, la metodología seguida parece adecuada para la descripción de los patrones de variación de los SNP realmente presentes en la población natural.

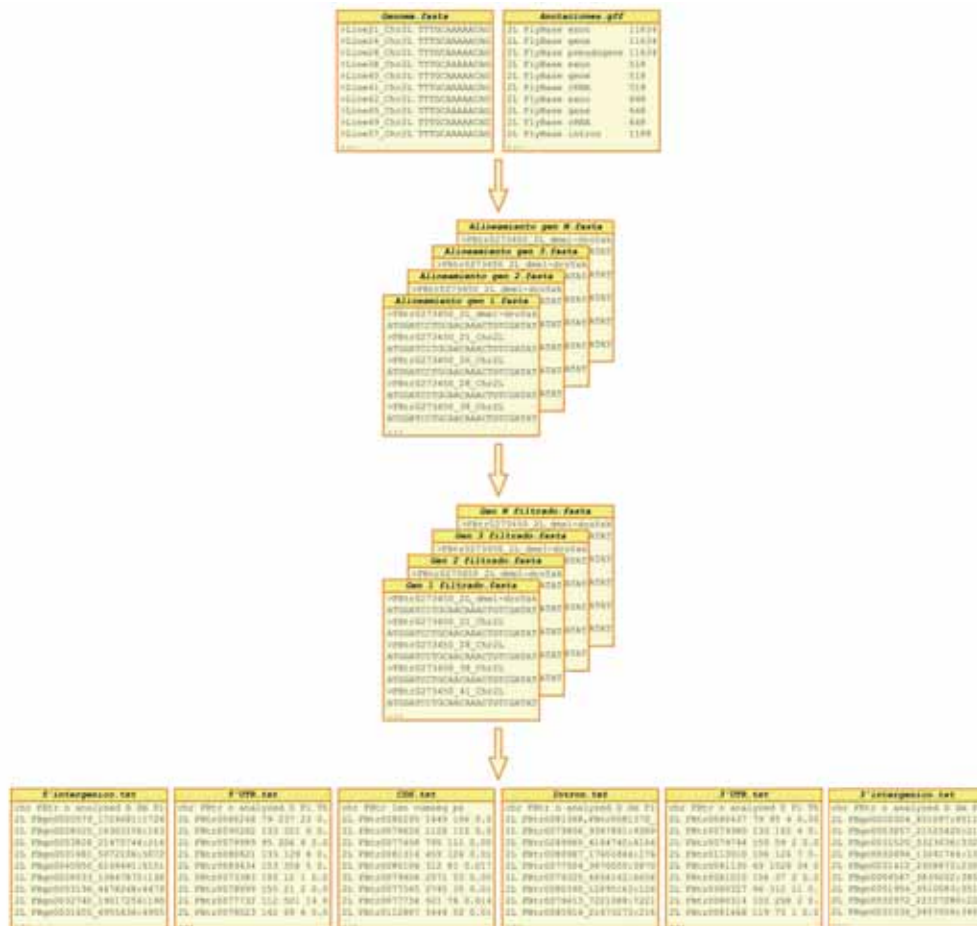
Para este trabajo se han utilizado 158 de las 168 líneas que se secuenciaron usando la tecnología de secuenciación de Illumina. Las lecturas provenientes de la resecuenciación se cartografiaron contra el genoma de referencia (versión 5.13 de *Flybase*). Son secuencias de alta calidad con una cobertura promedio de 21,4X por línea y con lecturas relativamente largas de 83pb en promedio (Anexo A). El SNP *calling* o imputación de los SNPs se realizó mediante un programa denominado JGIL (Stone *et al.* 2012), un método para identificar genotipos y variantes genéticas entre un gran grupo de cepas consanguíneas o isolíneas. Se han analizado los cromosomas 2, 3 y X. Para comprobar si había estructuración de la variación en la población se realizó un análisis filogenético de todas las líneas DGRP. El resultado fue un árbol en forma de estrella en el que cada una de las líneas se sitúa a una distancia filogenética similar a cualquier otra. Además, también se realizó un análisis de componentes principales (PCA, *Principal Component Analysis*) de la matriz de covarianza genética (Tabla suplementaria 2 en Mackay *et al.* 2012). Ambos análisis indican la práctica ausencia de estructura de la población en las líneas DGRP.

Uno de los problemas al analizar estas secuencias fue que se observó una mayor proporción de heterocigotos residuales de lo esperado ( $\sim 1\%$ ) en algunas líneas. Para evitar posibles sesgos en nuestro análisis se decidió eliminar estas posiciones (sección 3.2). En el Freeze 2 del DGRP (Huang *et al.* 2014), que se publicó posteriormente al análisis llevado a cabo en esta tesis, se amplió el número de líneas a 205 (820 autosomas) y se abordó el problema de los heterocigotos residuales. El 84% de las líneas tenían menos de un 2% de heterocigotos residuales, aproximadamente lo esperado teóricamente. De entre los 820 brazos autosómicos estudiados 62 contenían 9% o más de heterocigotos residuales. Sesenta de los 62 brazos cromosómicos contenían al menos una inversión heterocigótica mientras que los dos brazos restantes presentaban un cariotipo estándar. El exceso de sitios segregantes residuales en estos dos brazos podría explicarse suponiendo que una inversión segregaba en estas dos líneas cuando

se secuenció su genoma y que en el intervalo entre la secuenciación y la observación citológica para determinar el cariotipo se fijó la ordenación estándar. Se demuestra, por tanto, que hay una casi perfecta correlación entre las grandes regiones con heterocigotos residuales y la segregación de inversiones heterocigóticas. Luego, la decisión de eliminar estas posiciones resultó ser acertada.

## 2.2. Estimación de la diversidad y la divergencia

Los datos generados en esta tesis se basan en el alineamiento de las 158 secuencias para cada brazo cromosómico (DGRP) más el alineamiento de los taxones externos (véase sección 2.3). A lo largo de esta tesis se han utilizado dos aproximaciones para estudiar los patrones de variación nucleotídica: (i) una aproximación centrada en los genes codificadores (Figura 2.2), y (ii) una aproximación por ventanas consecutivas no solapantes a lo largo de los brazos cromosómicos.



**Figura 2.2.** Pipeline para la estimación de la diversidad nucleotídica y la divergencia en la aproximación centrada en genes

Con los alineamientos cromosómicos y las anotaciones génicas se obtienen los alineamientos génicos de cada unidad de análisis (fila 1-2). A continuación, cada alineamiento génico se filtra eliminando las líneas que contienen sitios ambiguos (Ns) (fila 3). Por último, se obtiene un archivo con las estimas de  $\pi$  y  $k$  para cada unidad de análisis. Cada línea de estos archivos contiene la información relativa a un alineamiento génico (fila 4).

---

### 2.2.1. Aproximación centrada en los genes codificadores

Esta aproximación se utilizó para realizar la primera parte de la tesis (secciones 3.1 y 3.2). La anotación de un gen codificador delimita una serie de regiones funcionales, codificadora (sinónima y no sinónima) y no codificadora (5' y 3' UTR, intrónica, 5' y 3' intergénica), cuyos patrones de variación y regímenes de selección se suponen distintos. Se desarrolló un conjunto de *scripts* que permitiera un procesamiento por etapas, *pipeline*, para calcular el polimorfismo y divergencia en cada uno de los fragmentos genómicos determinados a partir de las anotaciones disponibles (Figura 2.2). En el Anexo B se recoge el código del *script*. Los pasos seguidos se explican a continuación: i) En primer lugar, se realizó un *script* en Perl para fragmentar los alineamientos cromosómicos en alineamientos génicos en función de las anotaciones del genoma de referencia descargadas de *Flybase* (versión 5.13, <http://flybase.org>) (Anexo B). Es decir, la secuencia correspondiente a cada brazo cromosómico se “cortaba” siguiendo las anotaciones de cada clase de sitio generando las siguientes *unidades de análisis*: el conjunto de los exones (CDS), el conjunto de los intrones, el 5'UTR y el 3'UTR del transcrito más largo de cada gen y las regiones intergénicas. La unidad de análisis de cada región intergénica, la secuencias entre un gen y el siguiente, se separaron en dos mitades iguales, de modo que cada gen tiene dos regiones intergénicas asociadas: la región aguas arriba o 5' y la región aguas abajo o 3' del gen. ii) Una vez obtenidos estos alineamientos que denominamos génicos, se pasó a filtrar las secuencias en relación a los sitios ambiguos, aquellas posiciones en las que no se pudo determinar el nucleótido exacto en la secuencia, definidos con una N según la nomenclatura de la IUPAC (Anexo B). Si en una unidad de análisis alguna línea contenía uno o más sitios ambiguos (N) o posiciones heterocigóticas residuales, se eliminó la secuencia completa de esa línea. Así, cada unidad de análisis tiene su propio número de secuencias o líneas ( $n$ ), que suele diferir entre distintas unidades. La enorme cantidad de líneas facilitadas por el proyecto DGRP permitió realizar un filtrado tan exigente. Así, tras el filtrado, cada unidad de análisis contenía en promedio 108 líneas. A partir de estos alineamientos procesados se estimó el polimorfismo y la divergencia. iii) Se estimó la diversidad nucleotídica ( $\pi$ ) (Nei 1987) y el estimador de Watterson de la diversidad nucleotídica por sitio ( $\theta$ ) (Watterson 1975; Nei 1987) y la divergencia nucleotídica ( $k$ ) corregida por Jukes-Cantor (1969) para cada unidad de análisis utilizando *scripts* de Perl basados en *PDA2* (Casillas y Barbadilla 2006) (Anexo B: Programas Perl).

Además, se realizaron filtrados adicionales y específicos para la selección de estimas fiables y precisas de cada unidad de análisis. Éstas debían contener al menos 20 líneas sin ningún sitio ambiguo o heterocigoto residual. Tras este filtro se eliminaron 7,43% de las unidades de análisis. Para las secuencias codificadoras (CDS) la longitud de la secuencia ( $m$ ) debería ser superior a 50pb y no contener más de un codón stop. Sólo perdimos un 4,19% de

---

nuestros datos (4,06% por el criterio  $n < 20$ , un 0,02% por  $m < 50\text{pb}$  y un 0,1% por contener más de un codón stop). Para la clase de sitios UTR se añadió el criterio que la secuencia debería ser superior a 50pb (lo que supuso una reducción de global del 33,4%: un 0,92% por  $n < 20$ , un 8,34% por  $m < 50$  y un 24,14% por cumplir ambas condiciones) y para el resto de las regiones no codificadoras una  $m$  superior a 100pb. Ello supuso la reducción de un 50,23% de las unidades de análisis de la clase de sitios de intrones y un 40,19% de los intergénicos. Además, para las estimas de divergencia se analizaron únicamente los ortólogos con *D. yakuba* que presentaban 10% o menos de *gaps* entre los dos alineamientos (reducción adicional de un 13,11% CDS; un 19,88% UTR; un 36,82% intrones y un 38,64% intergénicas). Es decir, para las estimas de polimorfismo en total se analizó un 62,17% del genoma con un 95,82% (13019) de los CDS, un 66,60% (18100) de los UTR, un 49,77% (6763) de los intrones y un 47,11% (12804) de las regiones intergénicas. Para las estimas de divergencia se analizó un 47,36% del genoma con un 83,25% (11312) de los CDS, un 53,36% (14502) de los UTR, un 36,38% (4943) de los intrones y un 28,91% (7856) de las regiones intergénicas.

Los filtros aplicados redujeron de forma considerable el número de pb analizados a lo largo del genoma. Con este filtrado tan riguroso se pretendía priorizar la obtención de estimas precisas de calidad sobre la cantidad, reduciendo los errores que los valores extremos puedan introducir debido a artefactos por un pequeño número de líneas analizadas o regiones informativas demasiado cortas. La desventaja de un filtrado tan severo es que perdemos cierta información, por ejemplo, la de los intrones cortos ( $< 86\text{pb}$ ) cuya evolución parece ser distinta a la de los intrones largos (Haddrill *et al.* 2005a; Singh *et al.* 2009; Parsch *et al.* 2010; Clemente y Vogl 2012). Con este filtrado hemos podido analizar prácticamente el conjunto de todas las regiones codificadoras pero la mitad de las no codificadoras. Debido a la propia labilidad evolutiva de las regiones no codificadoras hay más posiciones con Ns en estas regiones, lo que incrementa la proporción de secuencias que no superan los filtros respecto a las secuencias de las regiones codificadoras. Por esta misma razón, la disminución es todavía más dramática para las estimas de divergencia puesto que nuestro análisis se ciñe a regiones ortólogas entre especies cercanas.

### 2.2.2. Aproximación por ventanas consecutivas no solapantes

El objetivo de esta aproximación es estudiar el efecto de variables genómicas tales como la recombinación, el desequilibrio de ligamiento, la región cromosómica o la densidad génica sobre los patrones de variación nucleotídica a lo largo de los brazos cromosómicos. Esta aproximación se utilizó para la segunda parte de la tesis (secciones 3.3 en adelante). Se definieron ventanas consecutivas no solapantes de 100kb y 10kb a lo largo del genoma. En este caso, en cada ventana los sitios codificadores se diferenciaron en: (1) los sitios 4-veces degenerados y (2) los sitios 0-veces degenerados; y los no codificadores en: (i) UTR, (ii)

---

intrones y (iii) regiones intergénicas. Se estimó el número de sitios segregantes (Nei, 1987), la diversidad nucleotídica ( $\pi$ ) (Nei, 1987) el estimador de Watterson de la diversidad nucleotídica por sitio ( $\theta$ ) (Nei, 1987) y la divergencia nucleotídica ( $k$ ) corregida por Jukes-Cantor (1969) para cada clase de sitio (estimaciones cedidas por M. Ràmia siguiendo el método, filtrado y control de calidad descritos en Mackay *et al.* 2012; Ràmia *et al.* 2012).

### 2.3. Elección del taxón externo

Se ha utilizado como taxón externo la especie *Drosophila yakuba*. El genoma de esta especie fue secuenciado por el consorcio de los 12 genomas (*Drosophila 12 Genome Consortium*, 2007) y su elección se debe a que este genoma presentaba la cobertura más alta (9.1X) de las especies que son filogenéticamente más cercanas a *Drosophila melanogaster* en el momento de comenzar esta tesis. El alineamiento de los genomas de *D. melanogaster* con *D. yakuba* en bloques (formato *axtNet*) se descargó del navegador genómico UCSC (<http://genome.ucsc.edu>). Los alineamientos en bloques se ordenaron y unieron a las secuencias alineadas del cromosoma completo de *Drosophila melanogaster* por Miquel Ràmia usando *scripts ad hoc*. Siempre se utilizaron las regiones ortólogas acorde con las anotaciones de *Flybase* (versión 5.13, <http://flybase.org>).

Además de por su mayor cobertura, el uso de la secuencia de *D. yakuba* parece ser adecuado para obtener estimaciones de selección adaptativa precisas gracias al tiempo de divergencia transcurrido entre las dos especies (Keightley y Eyre-Walker 2012). Si la divergencia es baja entre especies respecto a la variación nucleotídica dentro de las mismas, las estimaciones de la tasa de evolución adaptativa y la distribución de los efectos de la eficacia biológica pueden estar sesgadas debido a (i) una atribución errónea del polimorfismo a la divergencia, (ii) la contribución de polimorfismo ancestral a la divergencia, y (iii) las diferentes tasas de mutación neutra y adaptativa. Los autores proponen que la divergencia estimada a partir de especies estrechamente emparentadas (como es el caso de *D. melanogaster* con *D. simulans*) puede subestimar la tasa de evolución adaptativa en un ~10% o más. Por tanto, utilizar *D. yakuba* permite obtener estimaciones más fiables de la selección adaptativa.

### 2.4. Estimación de la Recombinación

En el momento de comenzar esta tesis no se disponía en *Drosophila melanogaster* de un mapa de recombinación unificado de alta resolución basado en una cantidad elevada de sitios segregantes cartografiados físicamente, como del que se dispone en humanos. Los datos de mapas genéticos disponibles en *flybase.org* son un compendio de información genética editado y rectificado repetidamente procedente de una gran cantidad de literatura de mapas genéticos, citogenéticos y físicos, de naturaleza muy heterogénea y en algunos casos conflictiva. En aquel momento las únicas estimaciones disponibles de la distribución de la tasa de entrecruzamientos por

---

pares de bases a lo largo del genoma eran las de Singh *et al.* (2005) actualizadas por Fiston-Lavier *et al.* (2010). Por tanto, en la primera parte de esta tesis se utilizó la recombinación a escala ancha como única medida de recombinación. La (1) **Tasa de recombinación a escala ancha** ( $C_{BS}$ , del inglés *Broad-Scale*) se obtuvo a través del servidor Web *Drosophila melanogaster Recombination Rate Calculator (RRC)* (Fiston-Lavier *et al.* 2010). Se estimó la tasa de recombinación en centiMorgan por megabase (cM/Mb) de cada unidad de análisis en la aproximación centrada en genes y en la aproximación por ventanas no solapantes en cada 100kb a lo largo del brazo cromosómico. La calculadora se basa en los mapas de Marey (Marais *et al.* 2001; Singh *et al.* 2005), donde las posiciones genéticas y físicas de 644 *loci* (marcadores moleculares) de la versión 5.19 del genoma de *D. melanogaster*, se ajustaron a una curva polinómica de tercer grado para cada brazo cromosómico. La tasa de recombinación se estimó como la derivada de la curva (Fiston-Lavier *et al.* 2010). Se utilizó el valor central de cada ventana como estima puntual de la recombinación de la ventana.

En el transcurso de esta tesis se publicaron una estimas de la tasa de recombinación mucho más precisas (Comeron *et al.* 2012). Por lo tanto, para los resultados de la segunda parte de esta tesis se añadieron estas nuevas estimas de recombinación. La (2) **Tasa de recombinación a escala fina** ( $C_{FS}$ , del inglés *Fine-Scale*) es una estima de alta resolución, la primera descripción integrada de la recombinación genómica de alta resolución obtenida en una especie. Los autores afirman que el mapa es 50 veces más detallado que el mapa de entrecruzamiento de alta resolución actual del genoma humano completo. Hemos utilizado la tasa de entrecruzamientos en ventanas no solapantes de 100kb que están disponibles en la dirección url <http://www.recombinome.com>. Los marcadores utilizados son SNPs y pequeños indels obtenidos de un subconjunto de 10 cepas de Carolina del Norte además de dos cepas africanas. El límite del tamaño de la ventana en nuestro análisis viene determinado por estas estimas de recombinación que sólo están disponibles en ventanas de 100kb.

Por último también nos propusimos estimar el (3) **parámetro de recombinación a escala poblacional** ( $\rho$ ). Hemos utilizado el paquete *LDhat* (versión 2.1; McVean *et al.* 2004) para estimar la tasa de recombinación poblacional a escala fina,  $\rho = 4N_e c$ , en ventanas no solapantes de 1kb, 10kb, 100kb y 1Mb.  $N_e$  es el *censo* efectivo de la población y  $c$  la tasa de recombinación por generación. Tan sólo se analizaron los sitios polimórficos con dos alelos (que representan el 97,2% de los SNPs) y se excluyeron del análisis aquellos sitios con más del 10% de sitios ambiguos (que representan el 19,9% de los SNPs). Se calculó la tasa de recombinación con el subprograma *interval* de *LDhat* usando los siguientes parámetros; -its 10.000.000, -bpen 5, -samp 5.000. Estos son valores recomendados para analizar grandes conjuntos de datos (McVean *et al.* 2004). Para las ventanas de 1Mb, 10kb y 1kb se modificó el primer parámetro (-its) a 1.000.000 iteraciones para reducir el coste computacional e intentar compensar el aumento de



---

tiempo debido al mayor número de ventanas a analizar siguiendo las instrucciones del manual del software. Hay disponibles varias tablas precomputadas (tablas de verosimilitud o *likelihood lookup table*) necesarias para ejecutar este programa (proporcionadas por McVean *et al.*; <http://ldhat.sourceforge.net/instructions.shtml>). Las tablas varían en el número de secuencias ( $n$ ) y el valor del parámetro poblacional de la tasa de mutación ( $\theta$ ). Usamos el subprograma *lkgen* para generar una nueva tabla de verosimilitud o *likelihood lookup table* de 158 secuencias a partir de una tabla precomputada de 192 secuencias y  $\theta = 0,001$  mediante el subprograma *lkgen*. Variaciones en el valor de  $\theta$  por sitio no parecen tener una influencia importante en las estimas de la tasa de recombinación (McVean *et al.* 2004). Los datos con una  $\rho > 0,06$  para las ventanas de 100kb (0,75% de los datos) y  $\rho > 0,1$  para las ventanas de 10kb (2% de los datos) fueron excluidos del análisis por considerarse valores extremos debido a errores metodológicos.

#### 2.4.1. Estimaciones corregidas de recombinación

La caracterización inicial de la recombinación en *Drosophila* (Morgan 1911) y más específicamente el descubrimiento de la ausencia de recombinación en los machos (Morgan 1912) se llevó a cabo en el laboratorio del Thomas Morgan (revisado en Ashburner *et al.* 2005). La implicación desde una perspectiva poblacional de que la recombinación se circunscribe a las hembras es que la mitad de los autosomas de la población experimentan sucesos de recombinación, mientras que en el cromosoma X esta proporción asciende a  $2/3$ . Por lo tanto, para comparar la recombinación autosomas *versus* el cromosoma X se deberían corregir las estimas de  $c$  por  $1/2$  en los autosomas y por  $2/3$  en el cromosoma X. Cuando se usa  $\rho$ , tanto el *censo* de población ( $N_e$ ) como la tasa de recombinación ( $c$ ) determinan su valor. Teóricamente, el cromosoma X tiene  $3/4$  el *censo* de población que los autosomas. Consecuentemente, para corregir nuestras estimas del parámetro de recombinación se debería multiplicar las estimas del cromosoma X por  $4/3$  y  $2/3$  y las estimas de los autosomas por  $1/2$ .

#### 2.5. Estimación de la densidad génica

Una de las variables estudiadas para intentar explicar la diversidad nucleotídica a lo largo del genoma (secciones 2.6 y 3.2) fue la densidad génica. Estas estimas fueron descargadas de *popDrowser* (Ràmia *et al.* 2012) para ventanas de 50kb. La densidad génica se define como el número de nucleótidos codificadores por ventana y se usan las anotaciones de la versión 5.13 de *Drosophila melanogaster* (*Flybase*).

#### 2.6. ANOVA de la variación nucleotídica y las variables genómicas

Se ha realizado un análisis de la varianza para probar los efectos que tienen sobre la diversidad nucleotídica la tasa de recombinación, la divergencia 4-veces degenerada, la densidad génica y la región genómica (centromérica *versus* no centromérica). En primer lugar se realizó una regresión lineal con los siguiente modelos: a) Para la aproximación centrada en

---

genes:  $\pi_{4\text{veces degenerado}} = f(\text{tasa de recombinación}, k_{4\text{veces degenerada}}, \text{región genómica})$ . b) Para la aproximación centrada en ventanas no solapantes:  $\pi_{4\text{veces degenerado}} = f(\text{tasa de recombinación}, k_{4\text{veces degenerada}}, \text{densidad génica}, \text{región genómica})$ . Después, se realizó un ANOVA de la regresión utilizando la suma de cuadrados tipo III (Fox 2008; Herr 1986; Langsrud 2003). Cuando los datos no son equilibrados, hay formas alternativas de calcular la suma de cuadrados en el ANOVA. Al menos hay tres aproximaciones, normalmente llamadas suma de cuadrados tipo I, tipo II y tipo III. Esta anotación, introducida originalmente por el paquete estadístico SAS, ahora se usa de un modo estandarizado (Langsrud 2003). El uso de uno u otro tipo de suma de cuadrados parece ser un tema controvertido en el campo de la estadística (Herr 1986), pero para el tipo de datos que teníamos en este análisis la suma de cuadrados tipo III es la aproximación que mejor permite analizar los datos (Sackton *et al.* 2009; Zahn 2010). El análisis se realizó para autosomas y el cromosoma X de forma separada.

## 2.7. Determinación del umbral de recombinación

Mackay *et al.* 2012 encontraron que la correlación positiva entre el polimorfismo y la tasa de recombinación en el nivel genómico desaparece cuando la recombinación local supera un umbral de aproximadamente 2 cM/Mb. En este trabajo damos un paso más usando un método estadístico más riguroso para determinar este umbral en el nivel cromosómico. Para realizar el análisis completo se utilizó el paquete estadístico *R* (<http://www.R-project.org>, 2008). El análisis de regresión es el método estadístico para describir la relación entre una variable dependiente y la variable independiente cuando ambas son continuas (Crawley 2007), como es nuestro caso. Construimos un modelo de regresión lineal en el que la variable dependiente es el polimorfismo corregido por la divergencia en las posiciones 4-veces degeneradas ( $\theta_{4\text{-veces}}/k_{4\text{-veces}}$ ) y la variable independiente la recombinación (ya sea  $c$  o  $\rho$ ).

A veces ocurre que la relación entre la variable dependiente e independiente no es lineal en todo el intervalo de definición de las variables, mostrando ciertos valores donde el efecto de la variable dependiente cambia de forma abrupta. Estos valores se denominan estadísticamente puntos de rotura (*breakpoints*). La palabra umbral también se puede utilizar, aunque algunos autores la usan sólo cuando no hay efecto en la variable dependiente antes del valor en cuestión (Muggeo 2003). En esta tesis se utilizarán de forma indistinta. Nótese que el significado de punto de rotura es estadístico y no tiene nada que ver con los puntos de rotura de las inversiones cromosómicas. Para determinar el punto de rotura entre el polimorfismo y la recombinación se utilizó un procedimiento de búsqueda iterativa mediante la función *breaks* (Crawley 2007). Se determina un posible valor umbral cuando un valor de recombinación distingue dos pendientes diferentes en el modelo de regresión según los intervalos de confianza: una pendiente positiva por debajo del valor umbral y una pendiente no diferente de cero por encima. De entre todos estos posibles valores umbrales, se elige el valor con el menor error cuadrático medio residual

---

como la mejor estima del valor umbral de recombinación. Seguimos este procedimiento para cada uno de los brazos cromosómicos y los autosomas en su conjunto para tres tamaños de ventana no solapante (1Mb, 100kb y 10kb).

Una vez determinados los valores umbrales de recombinación, se utilizó la función *piecewise* de R para estimar el modelo de regresión con dos segmentos. Se realizó un análisis de la varianza para contrastar las diferencias en la significación entre el modelos con y sin punto de rotura para comprobar cual se ajusta mejor a los datos. Además, debido a que la complejidad de los dos modelos es distinta, también se calculó el criterio de información de Akaike (AIC, Akaike 1974) para el modelo lineal (sin punto de rotura) o el modelo segmentado (con punto de rotura). AIC no proporciona una prueba acerca de la calidad de un modelo en el sentido de probar una hipótesis nula, sino que permite comprobar la calidad relativa de un modelo y por tanto seleccionar el que mejor se ajusta. AIC busca un balance entre la bondad de ajuste del modelo y su complejidad. Por tanto, es apropiado usarlo cuando los modelos a comparar son de complejidad distinta o están compuestos por un número distinto de variables.

Los valores umbrales de recombinación pueden interpretarse como el límite a *bloques de selección ligada*, es decir, a unidades discretas evolutivas del genoma que resumen la interacción histórica entre selección, mutación, deriva genética y recombinación que ha actuado sobre una región cromosómica.

## 2.8. Análisis estadístico

Para los análisis estadísticos adicionales también se utilizó el paquete estadístico R (<http://www.R-project.org>, 2008). Hemos realizado test no paramétricos porque los datos no se distribuyen en general de forma normal ni cumplen el criterio de homocedasticidad, incluso después de su transformación, debido principalmente a los valores igual o cercanos a cero que toman las variables usadas para medir la variación nucleotídica. El test de *Spearman* se ha utilizado para realizar las correlaciones. Para comparar dos muestras relacionadas se utilizó la prueba de *Wilcoxon* como alternativa a la prueba t de *Student*. Para probar si más de dos muestras siguen la misma distribución se utilizó el análisis no paramétrico de *Kruskal-Wallis*. Para probar si hay diferencias significativas entre distintas regiones del genoma (como los LSB y los NLSB, véase sección 3.3.3) para distintos estimadores, como  $r^2$  o la densidad génica, se realizó un remuestreo *bootstrap* de los datos 1.000 veces, seleccionando al azar cada ventana con sustitución (Canty y Ripley 2015; Davison y Hinkley 1997). Para algunos gráficos se utilizó la biblioteca *ggplot2* (Wickham 2009). Nótese que el paquete estadístico R utilizado presenta el límite de precisión de coma flotante en  $2,2 \times 10^{-16}$ , así que los valores de p de las pruebas de significación que son inferiores a este límite se representarán como 0.

---

## 2.9. Estimación del desequilibrio de ligamiento ( $r^2$ )

El desequilibrio de ligamiento es la asociación no aleatoria en dos o más *loci* en el genoma. Entre las medidas del desequilibrio de ligamiento se encuentran la  $D'$  (Lewontin, 1964) y la  $r^2$  (Hill y Robertson 1968) (Tabla 1.7).  $D'$  depende fuertemente del tamaño de la muestra y es más fácil que alcance su valor máximo si los haplotipos que se observan contienen alelos a bajas frecuencias. Aunque  $r^2$  también tiene sus limitaciones, debido a que es altamente dependiente de las frecuencias alélicas, es formalmente el coeficiente de correlación al cuadrado entre un par de *loci*, y creemos que es un estimador más intuitivo. Además, al ser una de las medidas del desequilibrio de ligamiento más utilizadas (Hartl y Clark 2007), nos permite la comparación de nuestros valores con estudios previos (Mackay *et al.* 2012; Langley *et al.* 2012). Se ha calculado  $r^2$  en distintos tamaños de ventanas utilizando el software *VariScan-2.0.3* (Hutter *et al.* 2006). El mínimo número de SNPs, incluyendo los *singletons*, utilizados para calcular  $r^2$  en una ventana es 2. Se compararon resultados a 100kb, 50kb, 10kb, 1kb, 500pb y 100pb. En ventanas de 100pb no había suficientes SNPs para calcular  $r^2$  en el 77,1% de las ventanas y en ventanas de 500pb en el 25,4% de las ventanas no se pudieron obtener estimas de  $r^2$ . Por tanto, descartamos el uso de estas estimas para estos pequeños tamaños de ventana. En ventanas de 1kb, este valor cae a un 11,7%, en ventanas de 10kb a un 2%, en ventanas de 50kb a 0,8% y en 100kb a 0,7%. En consecuencia, estos cuatro valores fueron los tamaños de ventana utilizados para el análisis.

## 2.10. Estimación de la razón $c/\mu$

La razón  $c/\mu$  es una estima relativa de la tasa de recombinación y la mutación por sitio nucleotídico. Es una medida útil para distinguir aquellas regiones en las que los niveles bajos de recombinación no son suficientes para separar las mutaciones que aparecen de forma independiente y cuyo destino hay que considerar de forma conjunta respecto a regiones donde las mutaciones no interfieren unas con otras (Lynch 2007). La razón  $c/\mu$  se estimó dividiendo las estimas de  $\theta$  de las posiciones 4-veces degeneradas por  $\rho$  en cada ventana. Asumimos que  $\theta$  de las posiciones 4-veces degeneradas son posiciones neutras y por tanto estas estimas son equiparables a  $4N_e\mu$ . Las estimas de  $\rho$  suponemos que son igual a  $4N_e c$ . Por tanto, mediante la simple división de ambas estimas obtenemos la razón  $c/\mu$  para cada ventana de 100kb. Estas estimas pueden estar afectadas por la selección, debido a que  $\rho$  puede estar subestimada en presencia de la selección ligada (Chan *et al.* 2012) y si la selección es eficiente, el valor de  $\theta$  de las posiciones 4-veces degeneradas puede verse afectado por la selección negativa, dando lugar a una subestima de la tasa de mutación real (Lawrie 2013). Es decir, que en las regiones donde la selección ligada es intensa, en los denominados bloques de selección ligada (LSB, véase sección 3.3.3), la estima de la razón  $c/\mu$  podría estar sesgada hacia valores menores debido a la subestima de  $\rho$ . No obstante, en las regiones donde la selección ligada no actúa, en los bloques

---

de selección no ligada (NLSB, véase sección 3.3.3), la eficiencia de selección podría ser mayor y la razón  $c/\mu$  se sobreestimaría debido a la subestima de  $\theta$  en las posiciones 4-veces degeneradas. Teniendo esto en cuenta, las diferencias observadas entre LSB y NLSB podrían ser menores que las observadas en esta tesis. Por esta razón se utilizan diferentes aproximaciones para evaluar la precisión de los bloques de selección ligada descritos en este trabajo.

## 2.11. Simulaciones del efecto de la recombinación

Debido a la importancia de validar los bloques de selección ligada (LSB, véase sección 3.3.3) estimados en este trabajo, quisimos inicialmente comprobar de forma teórica el efecto de la recombinación en la variación nucleotídica y en la eficiencia de la selección en un conjunto de secuencias. Para ello, realizamos simulaciones del efecto de la recombinación, ya sea el entrecruzamiento o la conversión génica, sobre un conjunto de secuencias de DNA sometidas a mutación, selección y deriva. Se ha utilizado el programa *Selection on Finite Sites under Complex Demographic Events (SFS\_CODE)* (Hernandez 2008) para este propósito. Este es un programa de simulación de genética de población que utiliza el modelo Wright-Fisher hacia delante (véase Cuadro 2 de la Introducción), siguiendo un modelo de sitios finitos con selección, recombinación (entrecruzamiento y/o conversión génica) y demografía para un número arbitrario de poblaciones. Es decir, que la población entera se sigue generación a generación desde el inicio de la simulación hasta el momento del muestreo a diferencia de lo que ocurre en simulaciones de coalescencia, donde la historia de una muestra se simula hacia atrás en el tiempo hasta su fundador. El programa genera por tanto una muestra de DNA, un alineamiento, en el que un sitio ha podido mutar o seleccionarse varias veces a lo largo de la simulación.

Los características y variables utilizadas para la simulación intentan imitar las observadas en datos reales y son las siguientes: i) la tasa de mutación = 0,01, ii) el *censo* de población inicial = 1000, iii)  $2N*50$  generaciones, es decir, 100.000 generaciones. iv) Los coeficientes de selección están basados en Schneider *et al.* 2011. Para simular las nuevas mutaciones deletéreas se supone que éstas siguen una distribución  $\gamma$  ( $4N_e\delta$ ) cuyo parámetro de forma (*shape parameter*) es 0,6. Además, 1000 es el efecto promedio de una mutación deletérea (*location parameter*). La fracción de nuevas mutaciones que son adaptativas es de 0,01 y la fuerza de la selección ( $4N_e s$ ) es igual a 10. v) La tasa de entrecruzamiento sigue una distribución de *Poisson* igual que la de conversión génica. Además, esta última tiene la media de la distribución en 500pb, que es el promedio observado en estudios genómicos experimentales (Comeron *et al.* 2012) (Cuadro 3). vi) Se han simulado 67kb de región codificadora en el que se supone que  $\frac{3}{4}$  de dichas posiciones son no sinónimas (~50kb).

---

## 2.12. Estimación de la selección

Se utilizó el programa *DEF-alpha* (Eyre-Walker y Keightley 2009) para estimar la fracción de sustituciones adaptativas ( $\alpha$ ) y la tasa relativa de sustituciones adaptativas respecto a la tasa de sustitución neutra ( $\omega_a$ , Schneider *et al.* 2011). Este método infiere la distribución del efecto de la eficacia biológica (DEF, *Distribution of Fitness Effect*) en los sitios funcionales (o putativamente afectados por la selección) a partir de espectro de frecuencias por posición (SFS, *Site Frequency Spectrum*). Uno de los principales problemas de las pruebas de selección basadas en SFS es que la selección ligada puede sesgar las estimas, subestimando  $\alpha$  o  $\omega_a$  (Messer y Petrov 2013). Recientemente se ha publicado que las estimas de  $\alpha$  del *DEF-alpha* son robustas, en cuanto a los efectos de la selección ligada, si se corrige el efecto de la demografía (utilizando la opción “2 epoch model” del programa; Messer y Petrov 2013). Por tanto, parece que el *DEF-alpha* es una aproximación adecuada para comparar regiones o bloques selectivos en los que la selección ligada es una fuerza principal (LSB: por debajo del valor umbral de recombinación) *versus* regiones donde el efecto de la selección ligada parece secundario (NLSB: por encima del valor umbral de recombinación) (véase sección 3.3.3).

*DEF-alpha* requiere estimas de divergencia y el SFS para las posiciones putativamente neutras (4-veces degeneradas) y selectivas (0-veces degeneradas). Se estimó el SFS para cada ventana no solapante de 100kb y 10kb, para los sitios 4 y 0-veces degenerados con *scripts ad-hoc* de Perl (cedidos amablemente por D. Castellano). Después, se ejecutó *DEF-alpha* para cada una de estas ventanas. Se utilizó una distribución  $\gamma$  para describir el efecto mutacional y se eliminó el polimorfismo que contribuye a la divergencia (Keightley y Eyre-Walker 2012).

## 2.13. Software utilizado

Para el cálculo del polimorfismo y divergencia en la aproximación centrada en genes codificadores se ha creado un *pipeline* en el que se integra el software *PDA2* (Casillas y Barbadilla 2006) (véase sección 2.2, el código del *pipeline* se recoge en ANEXO B). Para obtener las estimas centradas en ventanas continuas no solapantes se ha utilizado *popDrowser* y *VariScan 2* (véase sección 2.2; Mackay *et al.* 2012; Ràmia *et al.* 2012; Hutter *et al.* 2006). Las estimas de recombinación se han obtenido mediante i) el servidor web *Drosophila melanogaster Recombination Rate Calculator (RRC)* (Fiston-Lavier *et al.* 2010) y ii) el programa de *LDhat* (McVean *et al.* 2004) (véase sección 2.4). Las estimas del desequilibrio de ligamiento,  $r^2$ , se han efectuado utilizando *VariScan 2* (sección 2.9, Hutter *et al.* 2006). Las simulaciones del efecto de la recombinación sobre secuencias de DNA se han llevado a cabo con el software *SFS\_CODE* (sección 2.11; Hernandez 2008). Además, los valores de variables que miden la selección natural como  $\alpha$ ,  $\omega_a$  y  $\omega_d$  se estimaron con el programa *DEF-alpha* (sección 2.12; Eyre-Walker y Keightley 2009). Por último, el análisis estadístico se realizó con el paquete estadístico *R* (secciones 2.6, 2.7 y 2.8).

---

## Parte 3

# Resultados





---

## 3. Resultados

### 3.1. Descripción y caracterización de la variación en los genes codificadores de *Drosophila melanogaster*

A partir de los genomas completos de 158 individuos, se estimó la variación genética de cada gen y clase funcional dentro del gen (Materiales y Métodos sección 2.2). La estima de la diversidad nucleotídica global del genoma es de  $\pi = 0,0052$  y  $\theta = 0,0057$ . Los valores oscilan de la región más variable,  $\pi = 0,0143$  y  $\theta = 0,0129$ , en los sitios 4-veces degenerados del cromosoma 2L, a la región menos variable,  $\pi = 0,0008$  y  $\theta = 0,0013$ , en los sitios 0-veces degenerados del cromosoma X. La divergencia promedio al genoma ortólogo de *D. yakuba* es de  $k = 0,1474$ . Al igual que para el polimorfismo, las posiciones más variables son las 4-veces degeneradas del cromosoma 2L,  $k = 0,2506$ , sin embargo las posiciones menos variables son las 0-veces degeneradas del cromosoma 3L,  $k = 0,0423$  (Tabla 3.1; Tabla suplementaria 1). La variación del polimorfismo de una clase funcional a otra es mucho mayor (variaciones del menor al mayor de 10 veces para  $\theta$  y 18 veces para  $\pi$ , respectivamente) que para las estimas de divergencia (6 veces más variable el mayor respecto al menor). De hecho, el coeficiente de variación para las estimas de polimorfismo para todo el genoma es  $CV_{\pi} = 1,09$  y  $CV_{\theta} = 0,77$ , y para la divergencia es  $CV_k = 0,69$ .

La comparación de nuestras estimas con estudios previos de la misma población de *Drosophila melanogaster* (poblaciones norteamericanas) nos permite confirmar que la metodología utilizada (*SNP calling*, filtrado de los sitios ambiguos, etc, véase Materiales y métodos 2.1 a 2.3) es adecuada para el tipo de datos que tenemos. Hay que tener en cuenta, sin embargo, que los estudios previos disponibles abarcan un número mucho menor de tamaño de muestra,  $n$  de este estudio = 158 *versus*  $n = 6$  ó  $37$ , siendo  $n$  el número de líneas, y además no se analizó el genoma completo. Teniendo esto en cuenta, parece que las estimas observadas en esta tesis son comparables a las previamente descritas (Tabla 3.2).

**Tabla 3.1.** Polimorfismo y divergencia en el genoma de *D. melanogaster*

Clases funcionales	$\pi$						
	2L	2R	3L	3R	X	4/3X	Autosomas
<i>Total</i>	0,0064	0,0059	0,0055	0,0046	0,0031	0,0041	0,0055
<i>4-veces degenerados</i>	0,0143	0,0127	0,0123	0,0095	0,0071	0,0100	0,0120
<i>0-veces degenerados</i>	0,0018	0,0016	0,0015	0,0013	0,0008	0,0011	0,0016
<i>UTR</i>	0,0045	0,0054	0,0049	0,0042	0,0027	0,0034	0,0048
<i>Intrón</i>	0,0071	0,0092	0,0083	0,0076	0,0059	0,0060	0,0076
<i>Intergénico</i>	0,0070	0,0066	0,0062	0,005	0,0037	0,0048	0,0060

Clases funcionales	$\theta$						
	2L	2R	3L	3R	X	4/3X	Autosomas
<i>Total</i>	0,0069	0,0062	0,0062	0,0052	0,0036	0,0048	0,0060
<i>4-veces degenerados</i>	0,0129	0,0113	0,0115	0,0093	0,0069	0,0090	0,0111
<i>0-veces degenerados</i>	0,0024	0,0022	0,0022	0,0020	0,0013	0,0016	0,0022
<i>UTR</i>	0,0064	0,0057	0,0059	0,0051	0,0035	0,0045	0,0057
<i>Intrón</i>	0,0091	0,0080	0,0077	0,0063	0,0048	0,0053	0,0076
<i>Intergénico</i>	0,0077	0,0071	0,0070	0,0058	0,0044	0,0057	0,0068

Clases funcionales	$k$						
	2L	2R	3L	3R	X	Autosomas	
<i>Total</i>	0,1505	0,1431	0,1407	0,1452	0,1510	0,1505	
<i>4-veces degenerados</i>	0,2506	0,2467	0,2428	0,2423	0,2384	0,2506	
<i>0-veces degenerados</i>	0,0458	0,0457	0,0423	0,0428	0,0539	0,0458	
<i>UTR</i>	0,1219	0,1139	0,1141	0,1202	0,1320	0,1219	
<i>Intrón</i>	0,2323	0,2120	0,2186	0,2210	0,2220	0,2323	
<i>Intergénico</i>	0,1982	0,1889	0,1851	0,1866	0,1985	0,1982	

Estimas del promedio de la diversidad nucleotídica ( $\pi$  y  $\theta$ ) y de la divergencia ( $k$ ) para cada brazo cromosómico en cada clase funcional.

**Tabla 3.2.** Comparativa de las distintas estimas de la diversidad nucleotídica en poblaciones norteamericanas

<b>a) Sitios sinónimos o 4 veces degenerados</b>							
Estudio	Población	Sec.	Datos	# de individuos	Cromosoma	$\theta$	$\pi$
<i>Andolfatto 2001</i>	EU y NA	Sanger	82 loci	Pool	Autosomas	0,0155	0,0133
<i>Sackton et al. 2009</i>	NA	454	Genómicos	6	Autosomas	0,0103	-
<i>Mackay et al. 2012</i>	NA	Illumina	Genómicos	~100	Autosomas	0,0111	0,0120
<i>Langley et al. 2012</i>	NA	Illumina	Genómicos	37	Autosomas	0,0140	-
<i>Andolfatto 2001</i>	EU y NA	Sanger	82 loci	Pool	X	0,0198	0,0190
<i>Sackton et al. 2009</i>	NA	454	Genómicos	6	X	0,0069	-
<i>Mackay et al. 2012</i>	NA	Illumina	Genómicos	~100	X	0,0069	0,0071
<i>Langley et al. 2012</i>	NA	Illumina	Genómicos	37	X	0,0090	-

<b>b) Sitios no sinónimos o 0 veces degenerados</b>							
Estudio	Población	Sec.	Datos	# de individuos	Cromosoma	$\theta$	$\pi$
<i>Sackton et al. 2009</i>	NA	454	Genómicos	6	Autosomas	0,0012	-
<i>Mackay et al. 2012</i>	NA	Illumina	Genómicos	~100	Autosomas	0,0022	0,0016
<i>Langley et al. 2012</i>	NA	Illumina	Genómicos	37	Autosomas	0,0012	-
<i>Sackton et al. 2009</i>	NA	454	Genómicos	6	X	0,0008	-
<i>Mackay et al. 2012</i>	NA	Illumina	Genómicos	~100	X	0,0012	0,0008
<i>Langley et al. 2012</i>	NA	Illumina	Genómicos	37	X	0,0007	-

<b>c) Regiones intergénicas (*intron+intergénicas+UTR)</b>							
Estudio	Población	Sec.	Datos	# de individuos	Cromosoma	$\theta$	$\pi$
<i>Singh et al. 2007</i>	NA	Sanger	8 loci	~11	Autosomas	0,0094	-
<i>Sackton et al. 2009 (*)</i>	NA	454	Genómicos	6	Autosomas	0,0044	-
<i>Mackay et al. 2012</i>	NA	Illumina	Genómicos	~100	Autosomas	0,0068	0,0060
<i>Singh et al. 2007</i>	NA	Sanger	8 loci	~11	X	0,0085	-
<i>Sackton et al. 2009 (*)</i>	NA	454	Genómicos	6	X	0,0031	-
<i>Mackay et al. 2012</i>	NA	Illumina	Genómicos	~100	X	0,0044	0,0037

Comparación de nuestros datos (Mackay et al. 2012) con los promedios de las estimas de diversidad nucleotídica con otros estudios (previos y posteriores). Sec: Tipo de secuenciación, EU: europea, NA: norteamericana.

### 3.1.1. Diferencias entre clases funcionales de sitios

Con nuestros datos del genoma completo de decenas de individuos, hemos podido estimar, por primera vez, la diversidad nucleotídica de cada una de las clases funcionales en una población colonizadora, no Africana, de *Drosophila melanogaster*. El patrón de polimorfismo y

---

divergencia por clase de sitio es consistente entre cromosomas siguiendo el orden: sitios 4-veces degenerados > intrón > intergénico > UTR > sitios 0-veces degenerados (Figura 3.1, Tabla 3.1).

La posición relativa de la variabilidad de las clases de sitios coincide con estudios previos realizados con conjuntos de datos menores, exclusivamente con regiones del X, en poblaciones africanas (véase Tabla 3.3, Andolfatto 2005). Previo a este estudio existían indicios de que los intrones son más diversos que las regiones intergénicas (Halligan 2004) y que la divergencia sinónima es superior a la divergencia de los intrones y ésta a su vez a la divergencia de las regiones intergénicas (Haddrill *et al.* 2005a). Con nuestros resultados podríamos inferir que las posiciones más neutras del genoma son las posiciones 4-veces degeneradas puesto que sus niveles de divergencia y polimorfismo son los más elevados del genoma (Figura 3.1, Tabla 3.1). Las posiciones 4-veces degeneradas van seguidas de las posiciones intrónicas e intergénicas.

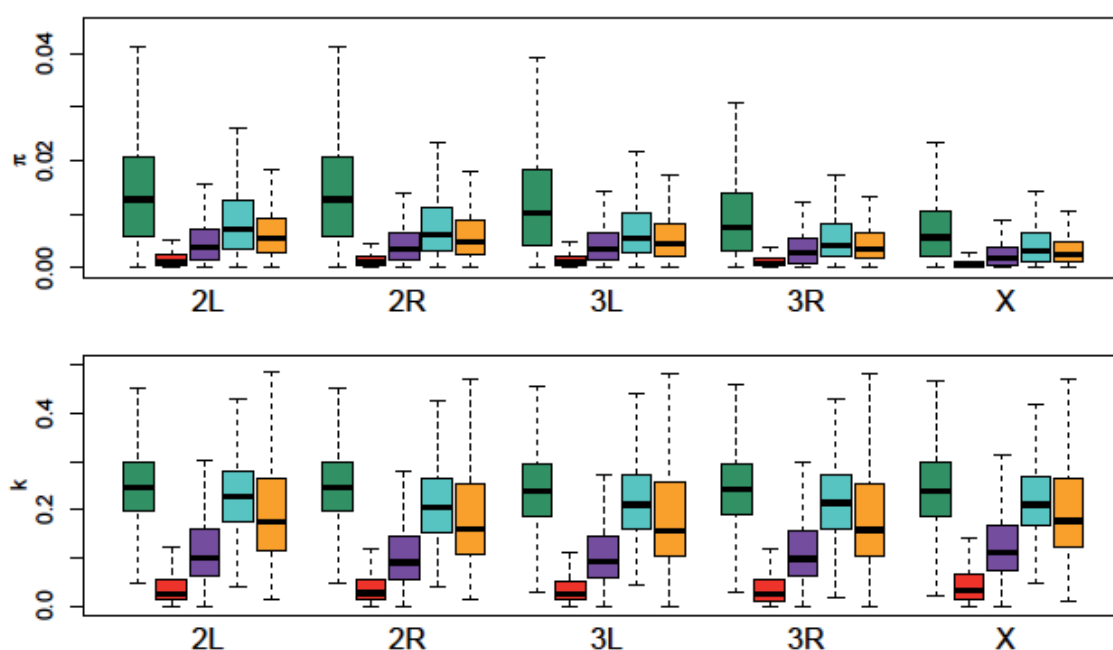
Nuestros datos parecen indicar que las regiones intergénicas y los intrones evolucionan de forma similar. Para determinar cuál de estas dos categorías es más variable se requeriría una subcategorización de estas clases funcionales, puesto que las diferencias son muy pequeñas y las muestras muy heterogéneas. Por ejemplo, hay evidencias de que los intrones difieren en sus patrones de polimorfismo en función de su longitud (Haddrill *et al.* 2005a; Singh *et al.* 2009; Parsch *et al.* 2010; Clemente y Vogl 2012) y deberían tratarse quizá como dos clases funcionales independientes (mayores y menores). Otro ejemplo son las regiones no-codificadoras altamente conservadas (CNSs), que difieren en sus niveles de diversidad de las regiones intergénicas no conservadas (no-CNSs) de *Drosophila melanogaster* (Casillas *et al.* 2007) y por tanto habría que analizarlas de forma separada. Además, en las estimas de divergencia los CNSs estarán sobrerrepresentados puesto que solo analizamos las regiones ortólogas.

La tasa de fijación de alelos en intrones y regiones intergénicas son muy similares a las posiciones 4-veces degeneradas, tan solo  $\sim 1,1$  y  $\sim 1,3$  veces menores, respectivamente. Sin embargo, hay menos alelos segregando en la población comparado con las posiciones 4-veces degeneradas,  $\sim 1,6$  veces menos en los intrones y  $\sim 2$  veces menos en las regiones intergénicas. Es decir, parece que la selección negativa es más intensa en estas clases de sitios previniendo que los alelos deletéreos segreguen en la población. En los UTR, la diversidad está todavía más constreñida,  $\sim 2,5$  veces menos que las posiciones 4-veces degeneradas y la divergencia también es  $\sim 2$  veces menor que las posiciones 4-veces degeneradas. Por último, las posiciones menos variables, que interpretamos como más constreñidas, parecen ser las posiciones 0-veces degeneradas, en las que la selección negativa está afectando muy substancialmente. Estas estimas son  $\sim 7,5$  veces menores que el polimorfismo 4-veces degenerado y  $\sim 5,6$  veces menores que la divergencia en las posiciones 4-veces degeneradas.

**Tabla 3.3.** Población norteamericana versus población africana

	Africana Cromosoma X	Norteamericana Cromosoma X	Norteamericana Todos los cromosomas
No sinónimas o 0 veces degeneradas	0,0018	0,0008	0,0014
UTR	0,0054	0,0027	0,0045
Regiones intergénicas	0,0111	0,0037	0,0056
Intrones	0,0125	0,0045	0,0071
Sinónimas o 4 veces degeneradas	0,0287	0,0071	0,0112

Comparación de los promedios por clase de sitio en el cromosoma X de un conjunto de poblaciones Africanas (Andolfatto 2005) y de la población norteamericana de *Drosophila melanogaster* estimadas en este estudio.



**Figura 3.1.** Polimorfismo y divergencia en el genoma de *D. melanogaster*

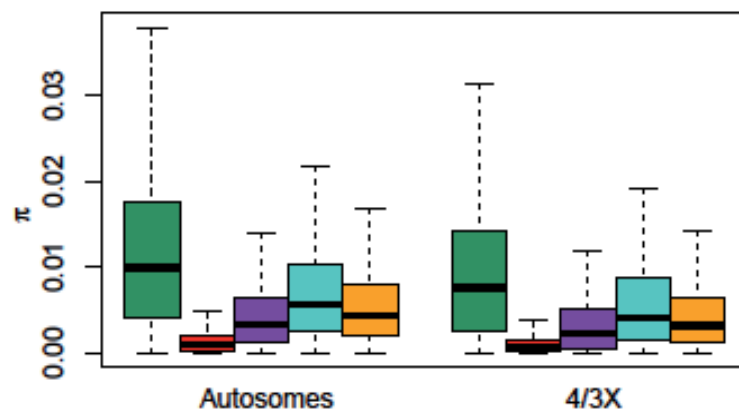
Polimorfismo,  $\pi$  (superior) y divergencia,  $k$  (inferior) para cada clase de sitio por brazo cromosómico. Clases de sitios; verde: posiciones 4 veces degeneradas, rojo: 0 veces degeneradas, morado: UTR, turquesa: intrónicas, amarillo: intergénicas.

### 3.1.2. Diferencias entre cromosomas

#### 3.1.2.1. Polimorfismo

Las estimas de diversidad nucleotídica difieren entre brazos cromosómicos con independencia de la clase funcional y siguen el orden siguiente, de mayor a menor polimorfismo,  $2L > 2R > 3L > 3R > X$  (*Kruskal-Wallis* valor  $p = 0$ ). Entre autosomas, se observan diferencias significativas entre brazos cromosómicos para cada clase funcional (0-veces degenerado: *Kruskal-Wallis* valor  $p = 0$ ; resto de clases funcionales: *Kruskal-Wallis* valor  $p = 0$ ). Como ya se había publicado previamente (Sackton *et al.* 2009), en la población norteamericana de *D. melanogaster* la variación en el cromosoma X se encuentra reducida

respecto a la diversidad autosómica en todas las clases de sitios (Figura 3.1, Tabla 3.1, *valor p* de la prueba de Wilcoxon = 0). Si la razón de los sexos es igual, se espera que la diversidad del cromosoma X sea  $\frac{3}{4}$  la de los autosomas, debido a la diferencia en el censo efectivo de la población. Para probar si la reducción observada en el X se desvía respecto a lo esperado, se han normalizado las estimas de diversidad del cromosoma X multiplicándolas por  $\frac{4}{3}$ . Igual que en estudios previos (Sackton *et al.* 2009), tras la normalización de la diversidad del cromosoma X, ésta sigue siendo significativamente inferior a la diversidad de los autosomas (Figura 3.2, Tabla 3.1; *valor p* de la prueba de Wilcoxon = 0). Además, esta reducción es consistente para todas las clases de sitios. Sin embargo, al comparar la variación del cromosoma X respecto a cada uno de los autosomas se observa la reducción del cromosoma X en todas las clases de sitios excepto para los sitios 4-veces degenerados (*valor p* de la prueba de Wilcoxon = 0,9187) y las regiones intergénicas (*valor p* de la prueba de Wilcoxon = 0,444) del cromosoma 3R (Figura 3.3, Tabla 3.1). Es decir, que al menos en las posiciones 4-veces degeneradas la diversidad del cromosoma X es similar a la de uno de los autosomas, el 3R.

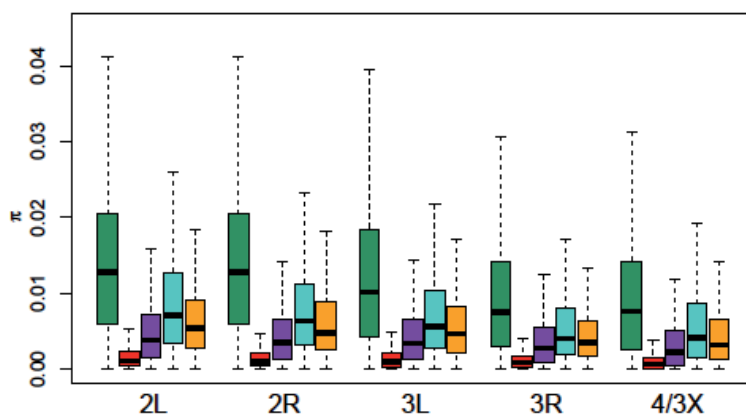


**Figura 3.2.** Polimorfismo genómico de autosomas versus polimorfismo genómico corregido para el cromosoma X

Comparación del polimorfismo,  $\pi$ , de los autosomas y  $\frac{4}{3}X$  para cada clase de sitio. Clases de sitios; verde: posiciones 4 veces degeneradas, rojo: 0 veces degeneradas, morado: UTR, turquesa: intrónicas, amarillo: intergénicas.

Se han propuesto varias hipótesis para explicar la disminución de los sitios sinónimos (en nuestro caso correspondería a los 4-veces degenerados) en el cromosoma X respecto a los autosomas. Las hipótesis propuestas para explicar la reducción en la variación sinónima del X respecto a los autosomas son (i) selectivas, (ii) demográficas y/o (iii) mutacionales (Begun y Whitley 2000; Wall *et al.* 2002; Singh *et al.* 2007; Pool y Nielsen 2008). (i) Si en promedio las nuevas mutaciones que son positivamente seleccionadas son a su vez parcialmente recesivas, los procesos de arrastre serán más eficientes en el cromosoma X porque estas mutaciones se exponen directamente al escrutinio de la selección en machos debido al estado hemigigótico del cromosoma X. Ello puede conllevar a una reducción de los sitios neutros ligados en el X

(Charlesworth *et al.* 1987, Orr y Betancourt 2001; Betancourt *et al.* 2004). (ii) En principio una desviación de la razón de los sexos a favor de los machos podría explicar este efecto porque reduciría el censo efectivo de la población del cromosoma X relativo a los autosomas, y esto disminuiría los niveles esperados de diversidad nucleotídica (Caballero 1995; Charlesworth 2001). Sin embargo, es improbable que esta desviación en la razón de los sexos explique por sí sola la reducción de la diversidad nucleotídica. Para producir la razón de diversidad X:A que observamos se necesitaría al menos el doble de machos que de hembras, lo que es improbable en ausencia de un impulso meiótico (*meiotic drive*) significativo que distorsione la razón de los sexos. Por otro lado, sí que se ha demostrado (Pool y Nielsen 2007, 2008) que cambios en la estructura de la población pueden producir la razón de diversidad que se observa en las poblaciones de *Drosophila*. Finalmente, (iii) una tasa de mutación distinta por sexo podría explicar la razón de diversidad X:A observada, aunque hasta la fecha no hay ninguna evidencia de la existencia de este sesgo en la tasa de mutación en *Drosophila* (Bauer y Aquadro 1997; Keightley *et al.* 2009; Schrider *et al.* 2013).



**Figura 3.3.** Polimorfismo cromosómico de autosomas versus polimorfismo cromosómico corregido para el cromosoma X

Comparación del polimorfismo,  $\pi$ , para cada clase de sitio y brazo cromosómico incluido el cromosoma X corregido por el esperado teórico. Clases de sitios; verde: posiciones 4 veces degeneradas, rojo: 0 veces degeneradas, morado: UTR, turquesa: intrónicas, amarillo: intergénicas.

Tras realizar el análisis cromosoma por cromosoma también tenemos que explicar la reducción de la variación en los sitios 4-veces degenerados del 3R respecto al resto de los autosomas (en ambos casos el *valor p* de la prueba de Wilcoxon = 0). Todas las explicaciones ofrecidas hasta el momento sólo son aplicables al cromosoma sexual, por lo tanto tendríamos que buscar hipótesis diferentes para la reducción en el cromosoma 3R. Una hipótesis plausible sería suponer que el mayor número de inversiones polimórficas que se han descrito en este autosoma (Corbett-Detig y Hartl 2012) podría reducir la recombinación y ésta a su vez la variación de los sitios neutros ligados a los selectivos. Sí así fuera, en este cromosoma la selección ligada sería más intensa.

---

Por otro lado, donde sí parece haber una diferencia clara entre autosomas y el X, incluso después de la corrección teórica debido a las diferencias en el censo efectivo, es en las clases de sitios donde la selección parece ser más intensa (como UTRs y sitios 0-veces degenerados, Tabla 3.1). Este efecto se podría explicar por procesos selectivos diferentes en el X respecto de los autosomas. Que en el X haya una reducción en la variación en los sitios no neutros indicaría que la eficiencia de selección en este cromosoma es superior al de los autosomas, reduciendo así los alelos ligeramente deletéreos de la población en los sitios como las posiciones 0-veces degeneradas o UTRs.

### 3.1.2.2. Divergencia

Las diferencias entre cromosomas para las estimas de divergencia no son tan notables como para la diversidad nucleotídica (véase Figura 3.1, Tabla 3.1). A pesar de ello, hay diferencias significativas entre cromosomas para cada clase funcional (*Kruskal-Wallis* valor  $p$  para 4-veces degeneradas = 0,0147; 0-veces degeneradas = 0; UTR = 0; intrones = 0; regiones intergénicas = 0) e incluso entre autosomas, excepto para los sitios 0-veces degenerados (*Kruskal-Wallis* valor  $p$  para 4-veces degeneradas = 0,0074; 0-veces degeneradas = 0,1999; UTR = 0; intrones = 0; regiones intergénicas = 0).

Si los sitios 4-veces degenerados se consideran neutros, la divergencia en esta clase de sitios se puede interpretar como una estima de la tasa de mutación neutra. El único cromosoma cuya divergencia es significativamente diferente para los sitios 4-veces degenerados es el cromosoma 2L (*valor p* de la prueba de Wilcoxon = 0,0017). Por tanto, no se observa el incremento de la divergencia neutra encontrada en trabajos anteriores en el cromosoma X en poblaciones africanas (Andolfatto 2005; Langley *et al.* 2012). Es decir, la mutación neutra no parece ser superior en el cromosoma X. Esto estaría de acuerdo con estudios experimentales de acumulación de mutaciones llevadas a cabo en 3 líneas europeas y 2 líneas norteamericanas (Keightley *et al.* 2009; Schrider *et al.* 2013). Además, en la clase funcional más claramente o fuertemente influenciada por la selección, los sitios 0-veces degenerados, hay diferencias significativas entre el X y los autosomas (*valor p* de la prueba de Wilcoxon = 0), pero no entre los autosomas (*Kruskal-Wallis* valor  $p$  = 0,1999), lo que corroboraría la hipótesis de que hay presiones selectivas diferentes entre el X y los autosomas. Que la  $k$  de los sitios 0-veces degenerados sea superior en el X (Tabla 3.1 y Tabla suplementaria 1) se podría interpretar de dos maneras opuestas. La  $k$  es mayor porque el X (i) experimenta una mayor eficiencia de selección y por tanto una mayor fijación de mutaciones adaptativas o (ii) experimenta una eficiencia de selección menor y se fijan un mayor número de alelos ligeramente deletéreos.



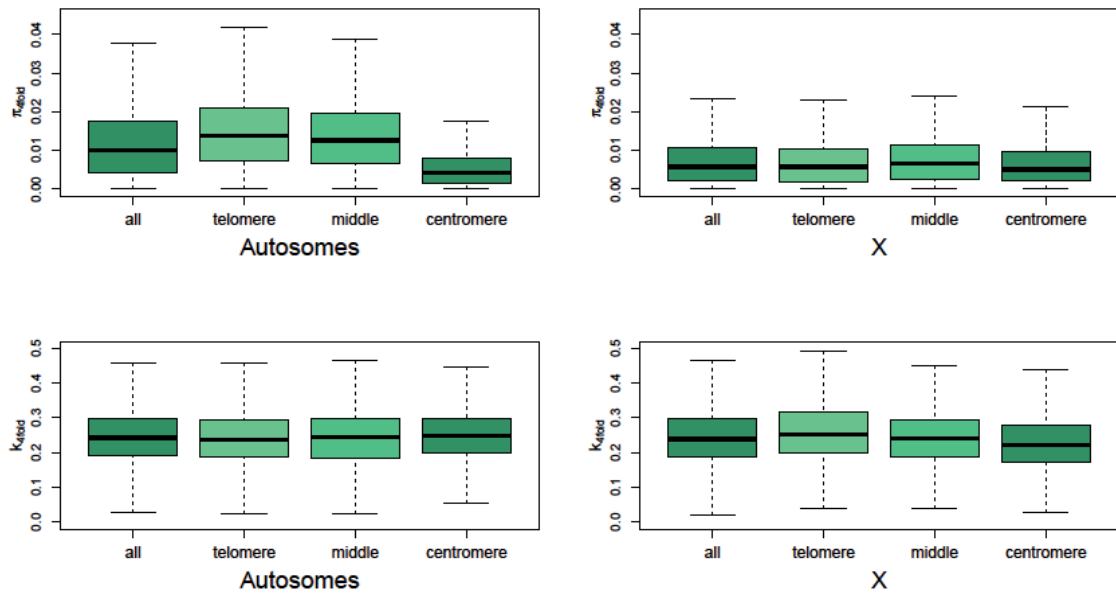
---

### 3.1.3. Diferencias entre regiones cromosómicas

Para analizar el potencial efecto de la región cromosómica sobre la variación nucleotídica cada cromosoma se dividió en 3 partes equitativas: la telomérica, el tercio del brazo cromosómico que incluye el telómero, la centromérica, el tercio que incluye el centrómero, y la región central. A pesar de las diferencias en los niveles de polimorfismo entre autosomas, el patrón de diversidad nucleotídica y de divergencia a lo largo del cromosoma es similar en todos ellos. Contrariamente, el cromosoma X presenta un patrón claramente diferenciado del resto. En concreto, los autosomas muestran una disminución de la variación y de la eficiencia de la selección en la región centromérica mientras que en el X estas diferencias no se observan (Figura 3.4, Figura 3.5, Figura 3.6, Figura 3.7, Tabla 3.4).

#### 3.1.3.1. Posiciones 4-veces degeneradas

En la región centromérica de los autosomas hay una gran disminución de la  $\pi$  en las posiciones 4-veces degeneradas respecto a las regiones no centroméricas (de un 68,70%, *valor p* de la prueba de Wilcoxon = 0, Tabla 3.4, Figura 3.4). Entre las regiones no centroméricas, la  $\pi_{4-veces\ degenerada}$  en la región central es ligeramente inferior (8,03%, *valor p* de la prueba de Wilcoxon = 0) a la región telomérica. Si suponemos que los sitios  $\pi_{4-veces\ degenerada}$  no están afectados por la selección, la disminución de la diversidad nucleotídica en esta clase de sitio se puede deber a: i) una disminución en la tasa de mutación o ii) la acción de la selección ligada. Debido a que la divergencia de los sitios 4-veces degenerados es ligeramente superior en la región centromérica que en las regiones no centroméricas (3,86%, *valor p* de la prueba de Wilcoxon = 0), la tasa de mutación neutra debería también incrementar en esta región si suponemos que los sitios 4-veces degenerados son neutros. Bajo la hipótesis de mutación variable esperaríamos una reducción en la tasa de mutación allá donde los niveles de polimorfismo sean menores. Por lo tanto, parece que la hipótesis de la selección ligada cobra mayor relevancia y ésta sería más intensa a medida que nos acercamos al centrómero, dejando una huella muy clara en la región centromérica de los autosomas. Es decir, que a medida que nos acercamos al centrómero o bien la selección es más intensa o bien los sitios están más ligados. Como se argumentará más adelante, la hipótesis de que los sitios están más ligados debido a una reducción en la recombinación parece ser la hipótesis más plausible.



**Figura 3.4.** Polimorfismo y divergencia de los sitios 4 veces degenerados

Polimorfismo,  $\pi$  (superior) y divergencia,  $k$  (inferior) en los sitios 4 veces degenerados de las tres regiones equitativas de los autosomas y del cromosoma X.

En el cromosoma X, sin embargo, no se observa el patrón previo. La región central muestra una  $\pi_{4-veces\ degenerada}$  ligeramente, pero significativamente, superior (16,92%, *valor p* de la prueba de Wilcoxon = 0,0015, Tabla 3.4) a las regiones de los extremos del cromosoma. La divergencia, sin embargo, incrementa a medida que nos alejamos del centrómero, es decir, la región central presenta una divergencia un 7,47% superior a la región centromérica (*valor p* de la prueba de Wilcoxon = 0,0005) y la región telomérica es a su vez un 5,45% superior a la región central (*valor p* de la prueba de Wilcoxon = 0,0057) (Figura 3.4, Tabla 3.4). Nótese que la reducción del polimorfismo en el centrómero es equivalente a la reducción de la tasa de mutación neutra ( $k_{4-veces-degenerada}$ ) en esta región (Tabla 3.4), por lo que no parece que sea necesario invocar la selección ligada para explicar la reducción del polimorfismo en la región centromérica del cromosoma X. Sin embargo, en la región telomérica se observa un ligero incremento de la tasa de mutación neutra ( $k_{4-veces-degenerada}$ ) (*valor p* de la prueba de Wilcoxon = 0,0057), por lo que la reducción de la diversidad en esta región sí podría deberse al efecto de la selección ligada. Según esto, se esperaría que el patrón de recombinación fuera distinto en el X respecto a los autosomas. Sabemos que los niveles de recombinación en el X son superiores a los de los autosomas pero la tendencia continua siendo la misma con una reducción paulatina hacia el centrómero y una brusca reducción en la región telomérica (Singh *et al.* 2005; Comeron *et al.* 2012).

**Tabla 3.4.** Polimorfismo y divergencia en autosomas y el cromosoma X

$\pi$	Autosomas			X		
	Centromérica	Central	Telomérica	Centromérica	Central	Telomérica
<i>4-veces degenerados</i>	0,0058	0,0141	0,0150	0,0065	0,0077	0,0069
<i>0-veces degenerados</i>	0,0011	0,0018	0,0017	0,0009	0,0008	0,0008
<i>UTR</i>	0,0030	0,0054	0,0057	0,0027	0,0027	0,0028
<i>Intrón</i>	0,0039	0,0090	0,0094	0,0042	0,0048	0,0044
<i>Intergénico</i>	0,0036	0,0069	0,0073	0,0036	0,0038	0,0036

$\theta$	Autosomas			X		
	Centromérica	Central	Telomérica	Centromérica	Central	Telomérica
<i>4-veces degenerados</i>	0,0060	0,0127	0,0136	0,0064	0,0079	0,0065
<i>0-veces degenerados</i>	0,0016	0,0024	0,0024	0,0012	0,0013	0,0013
<i>UTR</i>	0,0034	0,0055	0,0035	0,0033	0,0037	0,0034
<i>Intrón</i>	0,0044	0,0088	0,0092	0,0045	0,0053	0,0046
<i>Intergénico</i>	0,0045	0,0077	0,0079	0,0042	0,0049	0,0042

$k$	Autosomas			X		
	Centromérica	Central	Telomérica	Centromérica	Central	Telomérica
<i>4-veces degenerados</i>	0,2508	0,2457	0,2427	0,2313	0,2487	0,2632
<i>0-veces degenerados</i>	0,0456	0,0451	0,0418	0,0520	0,0547	0,0549
<i>UTR</i>	0,1280	0,1143	0,1126	0,1251	0,1309	0,1405
<i>Intrón</i>	0,2257	0,2254	0,2117	0,2207	0,2336	0,2110
<i>Intergénico</i>	0,2039	0,1916	0,1759	0,2093	0,2037	0,1818

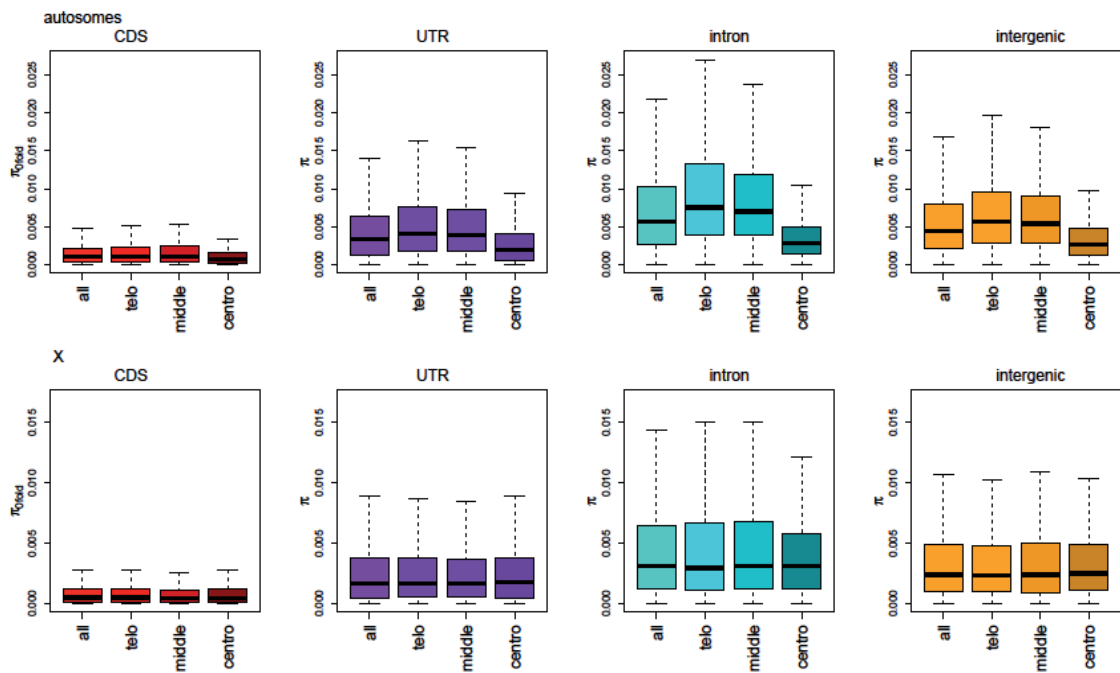
Estimas del promedio de la diversidad nucleotídica ( $\pi$  y  $\theta$ ) y de la divergencia ( $k$ ) para cada región cromosómico en cada clase funcional para los autosomas y el cromosoma X.

---

### 3.1.3.2. Posiciones putativamente funcionales

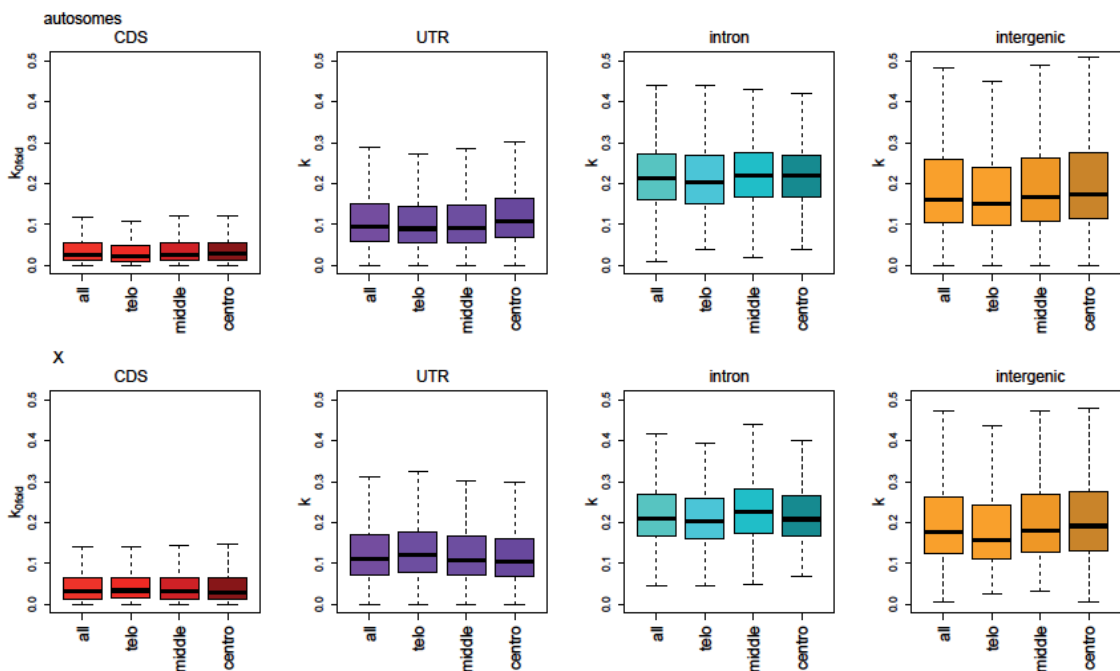
En los autosomas, al igual que para los sitios 4-veces degenerados, todas las clases de sitios tiene una menor diversidad en la región centromérica (0-veces degeneradas: 36,36%, UTR: 50%, intrón: 61,1%, intergénica: 50,91%, un  $\sim$ 49,6% en promedio para todas las clases de sitios conjuntamente, *valor p* de la prueba de Wilcoxon = 0). Cuando hay diferencias significativas entre las regiones no centroméricas, éstas son mucho más débiles y pequeñas (UTR: 4,88%, *valor p* de la prueba de Wilcoxon = 0,0076; intergénica: 3,57%, *valor p* de la prueba de Wilcoxon = 0,0125) (Figura 3.5, Tabla 3.4). Además, la regiones centroméricas muestran una *k* significativamente superior a la *k* de las regiones no centroméricas (0-veces degeneradas: 15,52% , *valor p* de la prueba de Wilcoxon = 0; UTR: 15,4%, *valor p* de la prueba de Wilcoxon = 0; intergénica: 9,39% , *valor p* de la prueba de Wilcoxon = 0). Esto es así excepto para las posiciones intrónicas, en las que sólo se observan diferencias significativas en la disminución de la región telomérica (7,85%, *valor p* de la prueba de Wilcoxon = 0) (Figura 3.6, Tabla 3.4). Dos hipótesis no excluyentes mutuamente podrían explicar estas observaciones: 1) los arrastres selectivos disminuyen el polimorfismo de forma global (no sólo para los sitios neutros) en las regiones centroméricas de los autosomas ó 2) la eficiencia de selección es mayor en esta región del genoma reduciendo el número de alelos ligeramente deletéreos.

En el cromosoma X, el patrón de polimorfismo es nuevamente diferente para todas las clases de sitios (Figura 3.5, Tabla 3.4), pero a diferencia de las posiciones 4-veces degeneradas, no hay diferencias significativas en las tres regiones del cromosoma. Para la divergencia, las únicas diferencias que se observan son entre las regiones teloméricas y el resto, siendo superior en el caso de los UTR (10,53%, *valor p* de la prueba de Wilcoxon = 0,0099) e inferior en el caso de las posiciones intergénicas (13,52%, *valor p* de la prueba de Wilcoxon = 0,0013; Figura 3.6). En general, todo apunta a que el efecto de la selección ligada es inferior en este cromosoma y no se observa disminución de la diversidad en las regiones donde la recombinación es menor como en el centrómero. A partir de estas observaciones no podríamos excluir la hipótesis de que la eficacia de la selección es equiparable en todo el genoma y que la fracción de los alelos ligeramente deletéreos es similar.



**Figura 3.5.** Polimorfismo en las distintas clases de sitios del genoma

Polimorfismo,  $\pi$ , para los distintos clases de sitios no neutros para las tres regiones equitativas de los autosomas y el cromosoma X. Clases de sitios; rojo: 0 veces degeneradas, morado: UTR, turquesa: intrónicas, amarillo: intergénicas.



**Figura 3.6.** Divergencia en las distintas clases de sitios del genoma

Divergencia,  $k$ , para los distintos clases de sitios no neutros para las tres regiones equitativas de los autosomas y el cromosoma X. Clases de sitios; rojo: 0 veces degeneradas, morado: UTR, turquesa: intrónicas, amarillo: intergénicas.

---

### 3.1.3.3. Razón $k_a/k_s$ y $\pi_a/\pi_s$

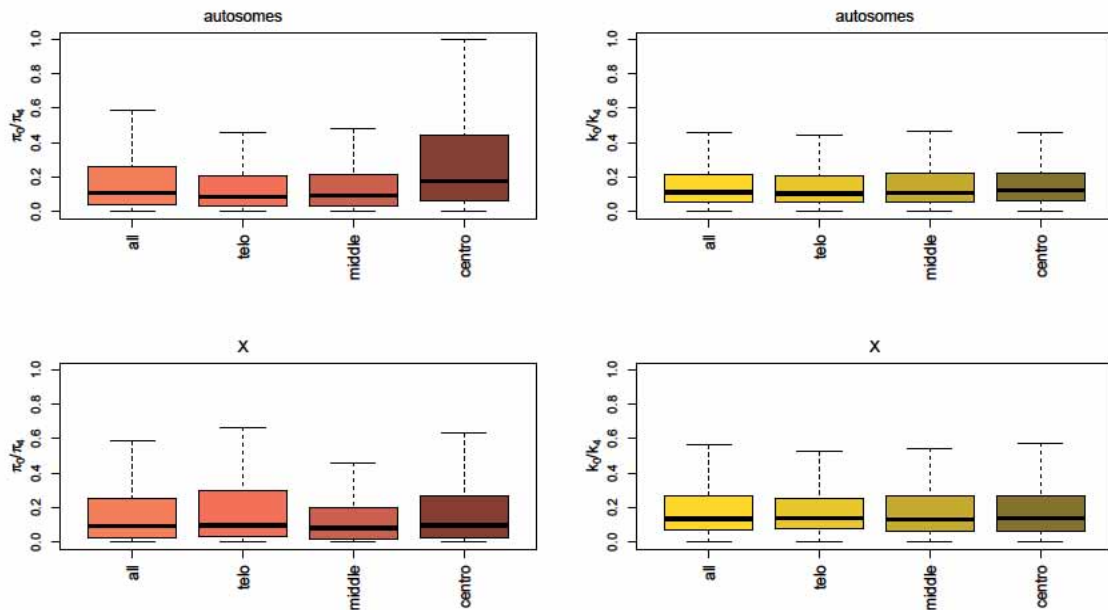
Las dos hipótesis para explicar la reducción del polimorfismo en la región centromérica de los autosomas son: (i) un incremento en la eficiencia de selección en esta región, o bien (ii) un efecto mayor de la selección ligada. Para tener una visión más completa de lo que puede estar sucediendo en las diferentes regiones cromosómicas, hemos comparado también la razón de los sitios 0-veces degenerado y los sitios 4-veces degenerados de la divergencia *versus* el polimorfismo:  $k_0/k_4$  *versus*  $\pi_0/\pi_4$  (Figura 3.7).

En los autosomas se observa un claro incremento de la razón  $\pi_0/\pi_4$  en las regiones centroméricas respecto a las no centroméricas (47,52%, *valor p* de la prueba de Wilcoxon = 0). Parece por tanto, que hay una disminución de la eficiencia de la selección en la región centromérica de los autosomas. Además, las regiones centroméricas de los autosomas muestran también una razón  $k_0/k_4$  ligeramente superior a las regiones no centroméricas (11,33%, *valor p* de la prueba de Wilcoxon = 0,0004). Los resultados previos parecen indicar que este incremento se debe principalmente a cambios en la divergencia de las posiciones 4-veces degeneradas. Si suponemos que estas posiciones son neutras, el incremento de la divergencia podría interpretarse como resultado de una tasa de mutación ligeramente superior en esta región. Bajo la hipótesis de que la recombinación es mutagénica esperaríamos una reducción de la tasa de mutación allí donde la recombinación sea reducida, pero no un incremento. Por tanto, la causa de la reducción en el polimorfismo en la región centromérica de los cromosomas parece debida a que la baja recombinación de la región incrementa el efecto de la selección ligada. Esto además estaría de acuerdo con lo propuesto previamente (Begun y Aquadro 1992; Hey y Kliman 2002; Presgraves 2005; Begun *et al.* 2007; Andolfatto 2007).

En el cromosoma X no hay diferencias significativas entre las regiones. El claro incremento de  $\pi_0/\pi_4$  en la región centromérica que muestran los autosomas no se observa pues en el X. La región central muestra una ligera disminución de la razón respecto de las regiones de los extremos de los cromosomas (16,32%, *valor p* de la prueba de Wilcoxon = 0,025, véase Figura 3.7). Por lo tanto, el cromosoma X muestra un patrón de variación muy diferente a los autosomas y se requieren análisis adicionales para comprender que ocurre en este cromosoma.

Este último análisis además nos permite distinguir entre las dos hipótesis barajadas a lo largo de este apartado. Por un lado, 1) la eficiencia de selección del X es menor (porque  $k_{0-veces-degenerado}$  es mayor en este cromosoma) y por tanto, la selección ligada debería ser más intensa que en los autosomas para observar la misma disminución de la diversidad. Bajo esta hipótesis, la región con menor recombinación, el centrómero, debería mostrar una mayor disminución en el X que no se observa. Por tanto, 2) parece que la eficiencia en el cromosoma X es mayor (porque  $k_{0-veces-degenerado}$  es mayor en este cromosoma), pero el efecto de la selección ligada es menor y por tanto no se observa un reducción del polimorfismo tan grande en la región

centromérica. Nuevamente, esta última hipótesis estaría de acuerdo con estudios previos (Baines *et al.* 2008). Por último, cabe reiterar que la tasa de mutación no parece ser la causa de las diferencias observadas entre cromosomas y entre las regiones cromosómicas. No hay diferencias significativas para  $k_4$ -veces-degenerado entre autosomas y X y, además,  $k_4$ -veces-degenerado parece ser ligeramente superior en las regiones centroméricas, que es donde se observa un menor polimorfismo, justo lo contrario de lo esperado bajo la hipótesis en la que la recombinación es mutagénica.



**Figura 3.7.** Razón  $\pi_0/\pi_4$  y  $k_0/k_4$  en distintas regiones del genoma

Razón  $\pi_0/\pi_4$  (rojo) y razón  $k_0/k_4$  (amarillo) para las tres regiones equitativas analizadas de los autosomas (superior) y el cromosoma X (inferior).

### 3.2. ANOVA para determinar las variables explicativas de la variación sinónima observada

Sabiendo que el polimorfismo neutro es  $\pi_s = 4N_e\mu$ , quisimos probar la importancia de variables relacionadas con el censo efectivo ( $N_e$ ) o con la tasa de mutación ( $\mu$ ) en la explicación del patrón de variación de la diversidad nucleotídica a lo largo del genoma. Una de las variables probadas es la divergencia de los sitios 4-veces degenerados, que puede interpretarse bajo la teoría neutra (Kimura 1983) como un indicador de la tasa de mutación neutra. Por otro lado, la correlación positiva consistentemente descrita en numerosos estudios entre la tasa de recombinación y el polimorfismo (revisado en Smukowski y Noor 2011; Cutter y Payseur 2013) puede deberse a que la recombinación es mutagénica o por su efecto sobre el censo efectivo de las distintas regiones de los cromosomas. Por lo tanto, hemos añadido la recombinación como variable independiente del ANOVA. Además, para la aproximación centrada en ventanas no solapantes hemos añadido la variable densidad génica como una medida indirecta de la densidad de dianas potenciales de selección natural en una región. Por último, hemos añadido la variable

---

región del genoma, ya sea centromérica o no centromérica, para medir el posible efecto de otras variables locales que no se hayan tenido en cuenta.

Para intentar determinar cuál de las variables previamente citadas explica mejor la variación observada realizamos un análisis de la varianza para la aproximación centrada en genes codificadores:  $\pi_{4\text{-veces degenerada}} = f(k_{4\text{-veces degenerada}}, \text{tasa de recombinación}, \text{región genómica})$  y para la aproximación centrada en ventanas no solapantes de 50kb:  $\pi_{4\text{-veces degenerada}} = f(k_{4\text{-veces degenerada}}, \text{tasa de recombinación}, \text{región genómica}, \text{densidad génica})$  (Tabla 3.5). En autosomas se observan diferencias muy significativas entre las variables analizadas. Es decir, que tanto la mutación neutra, como la recombinación, o como el contexto genético en el que se encuentra el gen, influyen significativamente en el patrón de variación neutro observado. Además, en el caso del análisis por ventanas observamos que la densidad génica tiene también cierta influencia en la variación observada. En el cromosoma X, sólo la recombinación y la región genómica afectan significativamente al patrón de polimorfismo. Ni la divergencia ni la densidad génica parecen tener un efecto en el cromosoma X. Los resultados muestran claramente que la tasa de recombinación local es la variable que mejor predice el polimorfismo neutro, tanto en los autosomas como en el cromosoma X. Además, se observa también un efecto significativo, aunque cuantitativamente menor, de la región genómica sobre el polimorfismo, tanto en autosomas como en el X. El efecto de la divergencia, todavía menor cuantitativamente, sólo se observa en los autosomas. Por último, la densidad génica es la variable que explica una menor fracción de la variación y sólo en los autosomas, pues en el X no tiene ningún efecto significativo. El efecto principal de la tasa recombinación y en menor medida de la divergencia ya había sido descrito previamente (Sackton *et al.* 2009), pero nuestro conjunto mayor de datos permite estimar con más precisión la importancia relativa de estas variables.

El ANOVA pone de manifiesto la importancia de la variable recombinación a la hora de explicar los patrones de variación a lo largo de los cromosomas. Este efecto se observa tanto en los autosomas como en el cromosoma X. Luego, a pesar del patrón de polimorfismo diferente observado en los autosomas respecto del cromosoma X (sección 3.1.3), la recombinación sigue siendo con mucho el factor principal para explicar la diversidad observada en este cromosoma. La hipótesis principal para dar cuenta de este efecto de la recombinación es la de la selección ligada (Begun y Aquadro 1992). Sin embargo, la hipótesis alternativa, que la recombinación sea mutagénica, no puede ser completamente descartada a partir de este análisis. La tasa de mutación, aunque con un efecto menor y secundario, podría también influir en los patrones de variación neutra observados en los autosomas.



**Tabla 3.5.** ANOVA del polimorfismo sinónimo y variables genómicas como la recombinación y mutación

<b>a) Aproximación centrada en genes</b>					
	<b>Suma de cuadrados</b>	<b>gl</b>	<b>F-valor</b>	<b>valor p</b>	
<b>Autosomas</b>					
<i>(Intercepto)</i>	0,00040	1	6,32	$< 2,2 \times 10^{-16}$	***
<i>Recombinación</i>	0,03190	1	499,90	$< 2,2 \times 10^{-16}$	***
<i>Región</i>	0,00823	1	129,02	$< 2,2 \times 10^{-16}$	***
<i>Divergencia</i>	0,00632	1	99,04	$< 2,2 \times 10^{-16}$	***
<i>Residuos</i>	0,52948	8298			
<b>X</b>					
<i>(Intercepto)</i>	0,00010	1	46,28	$1,462 \times 10^{-11}$	*
<i>Recombinación</i>	0,00066	1	69,01	$< 2,2 \times 10^{-16}$	***
<i>Región</i>	0,00008	1	4,08	0,0435	*
<i>Divergencia</i>	0,00002	1	0,44	0,5094	
<i>Residuos</i>	0,00732	415			
<b>b) Aproximación centrada en ventanas</b>					
	<b>Suma de cuadrados</b>	<b>gl</b>	<b>F-valor</b>	<b>valor p</b>	
<b>Autosomas</b>					
<i>(Intercepto)</i>	0,00275	1	86,07	$< 2,2 \times 10^{-16}$	***
<i>Recombinación</i>	0,00637	1	199,03	$< 2,2 \times 10^{-16}$	***
<i>Región</i>	0,00340	1	106,21	$< 2,2 \times 10^{-16}$	***
<i>Divergencia</i>	0,00224	1	69,88	$< 2,2 \times 10^{-16}$	***
<i>Densidad génica</i>	0,00027	1	8,43	0,0037	**
<i>Residuos</i>	0,05914	1848			
<b>X</b>					
<i>(Intercepto)</i>	0,00010	1	5,71	0,0173	*
<i>Recombinación</i>	0,00066	1	37,32	$5,844 \times 10^{-15}$	***
<i>Región</i>	0,00008	1	4,8	0,0290	*
<i>Divergencia</i>	0,00002	1	0,87	0,3522	
<i>Densidad génica</i>	0,00001	1	0,61	0,4364	
<i>Residuos</i>	0,00732	415			

(a) Anova de la regresión lineal del modelo  $\pi_{4fold} = \text{recombinación} + k_{4fold} + \text{región del genoma, centromérica o no centromérica}$ . Aproximación centrada en genes codificadores. (b) Anova de la regresión lineal del modelo  $\pi_{4fold} = \text{recombinación} + k_{4fold} + \text{región del genoma} + \text{densidad génica}$ . Aproximación centrada en ventanas no solapantes de 50kb.

---

### 3.3. Valor umbral de recombinación

Dos resultados previos nos movieron a profundizar en la relación entre el polimorfismo y la recombinación: (1) La recombinación parece ser la variable principal para explicar la variación neutra a lo largo del genoma (Sackton *et al.* 2009; Mackay *et al.* 2012). (2) A partir de un valor dado de recombinación estimado alrededor de 2 *cM/Mb*, la correlación entre el polimorfismo y la recombinación desaparece (Mackay *et al.* 2012). A partir de estos resultados, quisimos estudiar de forma más detallada la existencia de un posible valor umbral de recombinación y la relación entre el polimorfismo y la recombinación en los distintos cromosomas. Para ello, primero se actualizaron las estimas de recombinación clásicas ( $c_{FL}$ , Fiston-Lavier *et al.* 2010) utilizadas hasta el momento por medidas recientes de recombinación que apartaran una resolución mucho mayor, como son las estimas de tasa de recombinación obtenidas para *Drosophila melanogaster* que utilizan SNPs como marcadores en lugar de genes ( $c_c$ , Comeron *et al.* 2012) y estimas del parámetro de recombinación calculadas para esta población mediante el programa *LDhat* ( $\rho$ , McVean *et al.* 2004). Los detalles de cómo se estimaron estas nuevas métricas de recombinación se recogen en el apartado 2.4 de Materiales y Métodos.

#### 3.3.1. Comparación y validación de las estimas de recombinación en ventanas de 100kb

El patrón general de las tres medidas de recombinación ( $c_{FL}$ ,  $c_c$ ,  $\rho$ ) a lo largo del genoma en ventanas de 100kb es similar, exhibiendo un descenso más o menos uniforme conforme se va hacia los centrómeros y una caída drástica de la recombinación en la proximidad de los telómeros (Figura 3.8). A modo de confirmación, la correlación entre estas tres medidas es alta y muy significativa (*Spearman R*: 0,65; valor  $p = 0$ ) en todos los brazos cromosómicos (Tabla 3.6). En ventanas de menor resolución (1Mb), este patrón similar entre las tres estimas de recombinación es todavía más claro (Figura 3.8) y la correlación más fuerte (*Spearman R*: 0,80; valor  $p = 0$ , Tabla 3.6). Los valores promedio por brazo cromosómico de  $c_{FL}$  (2,295) y  $c_c$  (2,322) son similares entre ellos (Tabla 3.7). Los valores promedio por brazo cromosómico del parámetro de recombinación,  $\rho$ , no son comparables a la estimas de la tasa de recombinación, pues  $\rho$  es función de la tasa de recombinación y del censo efectivo de la población ( $\rho = 4N_e c$ ). Chan y colaboradores (2012) estiman el parámetro de recombinación promedio por brazo cromosómico para una subconjunto de datos (37 líneas) de esta misma población norteamericana de *D. melanogaster* utilizando un software distinto al utilizado en este trabajo (*LDhelmet*). Este software se basa en la misma metodología que el *LDhat*, pero supuestamente está orientado hacia las características del genoma de *Drosophila* en vez de al de humanos. Puesto que la publicación y disponibilidad de este programa fue posterior al cálculo de las estimas de  $\rho$  en esta tesis, no pudimos emplearlo. A pesar de ello, los resultados de  $\rho$  por kb son

comparables a los nuestros para los autosomas ( $A_{chan}$ : 12,42;  $A_{esta\ tesis}$ : 10,07) y el cromosoma X ( $X_{chan}$ : 14,8;  $X_{esta\ tesis}$ : 15,25). Esta similitud refuerza la validez de nuestras estimas de  $\rho$ .

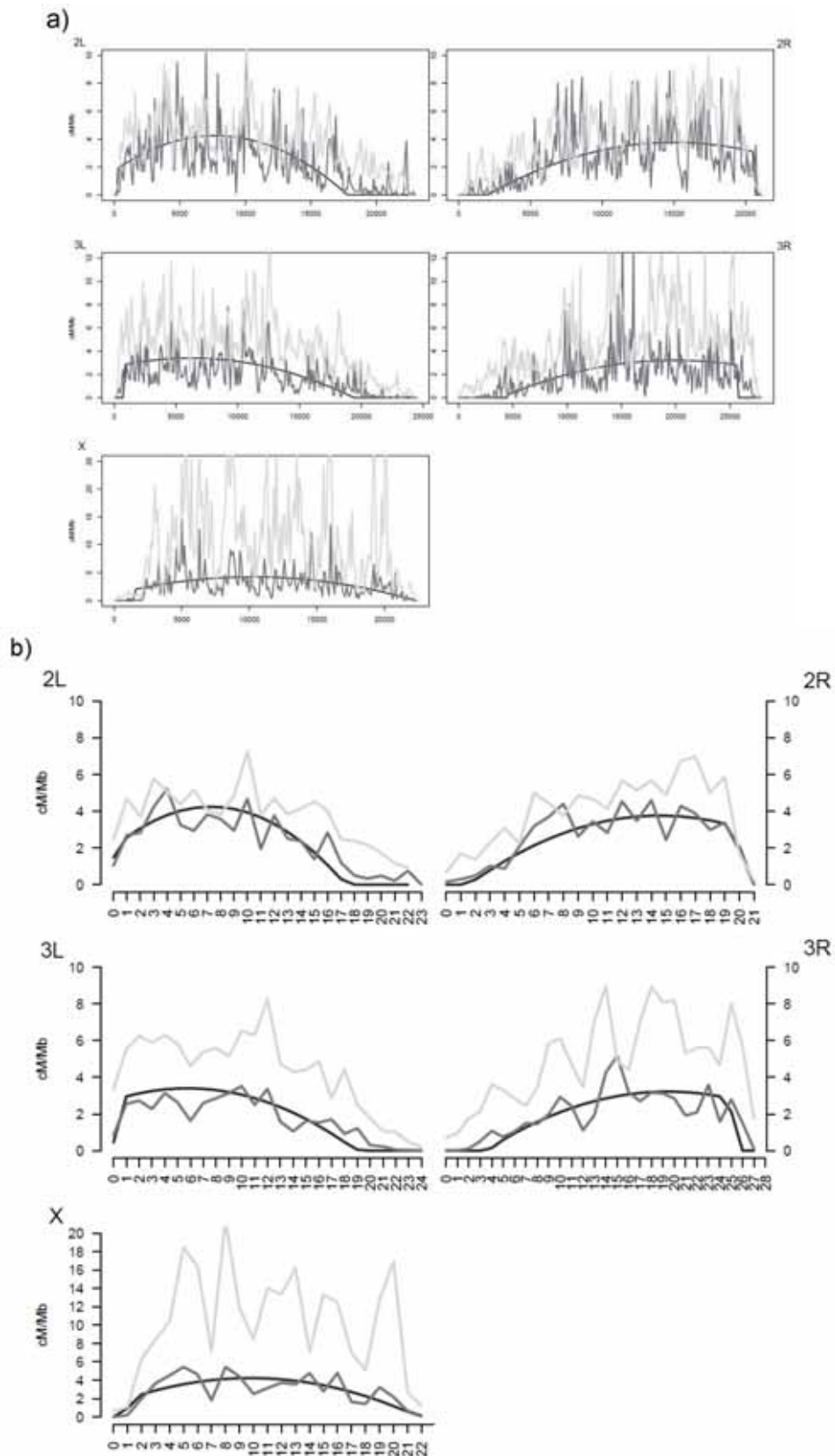
**Tabla 3.6.** Correlación entre las distintas estimas de recombinación a lo largo del genoma

100kb	$c_{FL}$ vs $c_c$		$c_{FL}$ vs $\rho$		$c_c$ vs $\rho$	
	R	valor p	R	valor p	R	valor p
TOTAL	0,6463	$<2,2 \cdot 10^{-16}$	0,6216	$<2,2 \cdot 10^{-16}$	0,6723	$<2,2 \cdot 10^{-16}$
2L	0,6274	$<2,2 \cdot 10^{-16}$	0,5853	$<2,2 \cdot 10^{-16}$	0,6446	$<2,2 \cdot 10^{-16}$
2R	0,6123	$<2,2 \cdot 10^{-16}$	0,6971	$<2,2 \cdot 10^{-16}$	0,6753	$<2,2 \cdot 10^{-16}$
3L	0,6554	$<2,2 \cdot 10^{-16}$	0,6483	$<2,2 \cdot 10^{-16}$	0,7414	$<2,2 \cdot 10^{-16}$
3R	0,6354	$<2,2 \cdot 10^{-16}$	0,6394	$<2,2 \cdot 10^{-16}$	0,7120	$<2,2 \cdot 10^{-16}$
X	0,4883	$1,204 \cdot 10^{-14}$	0,5396	$<2,2 \cdot 10^{-16}$	0,6417	$<2,2 \cdot 10^{-16}$

1Mb	$c_{FL}$ vs $c_c$		$c_{FL}$ vs $\rho$		$c_c$ vs $\rho$	
	R	valor p	R	valor p	R	valor p
TOTAL	0,8172	$<2,2 \cdot 10^{-16}$	0,7602	$<2,2 \cdot 10^{-16}$	0,8130	$<2,2 \cdot 10^{-16}$
2L	0,8461	$3,66 \cdot 10^{-7}$	0,7110	$1,43 \cdot 10^{-4}$	0,8300	$1,84 \cdot 10^{-6}$
2R	0,7807	$1,81 \cdot 10^{-5}$	0,8768	$8,68 \cdot 10^{-8}$	0,8216	$2,45 \cdot 10^{-6}$
3L	0,8178	$5,99 \cdot 10^{-7}$	0,7884	$2,86 \cdot 10^{-6}$	0,9054	$2,01 \cdot 10^{-6}$
3R	0,7969	$3,88 \cdot 10^{-7}$	0,7238	$1,34 \cdot 10^{-5}$	0,8139	$1,61 \cdot 10^{-6}$
X	0,6156	$2,19 \cdot 10^{-3}$	0,5879	$3,76 \cdot 10^{-3}$	0,7431	$7,56 \cdot 10^{-5}$

Correlación de Spearman (R) y *valor p* entre las 3 medidas de recombinación recolectadas y estimadas ( $c_{FL}$  = estimas clásicas de recombinación,  $c_c$  = estimas de recombinación de alta resolución y  $\rho$ , parámetro de recombinación histórica) para cada brazo cromosómico y el genoma completo (Total) en ventanas de 100kb y 1Mb.

Teniendo en cuenta la alta correlación, el patrón genómico semejante y la media similar por brazo cromosómico que muestran  $c_{FL}$  y  $c_c$ , elegimos  $c_c$  como tasa de recombinación (que denominaremos simplemente  $c$ ) debido a su mayor resolución. Por tanto, para el resto de esta tesis usaremos  $\rho$  como medida de recombinación histórica y  $c$  como medida de la tasa de recombinación actual (basado en los mapas genéticos cartografiados por Comeron *et al.* 2012). Las principales diferencias en el valor promedio de la tasa de recombinación entre los brazos cromosómicos se observan entre los autosomas y el cromosoma X (Tabla 3.7;  $c$ : *valor p* de la prueba de Wilcoxon = 0,  $\rho$ : *valor p* de la prueba de Wilcoxon = 0), como ya se había observado previamente con las estimas de recombinación clásicas y para el parámetro de recombinación  $\rho$  (Mackay *et al.* 2012; Langley *et al.* 2012; Chan *et al.* 2012).



**Figura 3.8.** Patrón de recombinación a lo largo de cada brazo cromosómico

Patrón de recombinación a lo largo del cromosoma en ventanas de a) 100kb y b) 1Mb. En negro son las estimas de la tasa de recombinación de Fiston-Lavier et al. 2010, en gris oscuro las estimas de la tasa de recombinación de Comeron et al. 2012 y en gris claro las estimas del parámetro de recombinación estimadas en esta tesis usando el software de McVean et al. 2004.

**Tabla 3.7.** Descripción estadística de las distintas estimas de recombinación por brazo cromosómico

Chr	$c_{FL}$				$c_c$				$\rho$			
	$\bar{x}$	$P_{50}$	$\sigma$	CV	$\bar{x}$	$P_{50}$	$\sigma$	CV	$\bar{x}$	$P_{50}$	$\sigma$	CV
<i>T</i>	2,295	2,850	1,436	0,626	2,322	1,876	2,124	0,808	11,040	9,802	7,558	0,685
<i>2L</i>	2,426	2,928	1,612	0,664	2,395	1,994	2,096	0,875	10,810	10,320	5,493	0,508
<i>2R</i>	2,481	3,140	1,345	0,542	2,662	2,386	2,142	0,805	9,894	9,786	5,542	0,560
<i>3L</i>	1,939	2,490	1,369	0,706	1,792	1,454	1,535	0,857	10,670	10,420	6,326	0,593
<i>3R</i>	1,872	2,375	1,287	0,688	1,963	1,523	1,993	1,015	9,083	7,880	5,690	0,626
<i>X</i>	2,912	3,335	1,313	0,451	2,954	2,441	2,587	0,876	15,260	12,220	11,717	0,768
<i>Autosomas</i>	2,178	2,733	1,403	0,650	2,179	1,741	1,977	0,907	10,070	9,516	5,819	0,578

Media ( $\bar{x}$ ), mediana ( $P_{50}$ ), desviación estándar ( $\sigma$ ) y coeficiente de variación ( $CV$ ) de las tres medidas de recombinación ( $c_{FL}$  = estimas clásicas de recombinación,  $c_c$  = estimas de recombinación de alta resolución y  $\rho$ ) en ventanas de 100kb para los distintos cromosomas y el genoma completo (T=total).

### 3.3.2. Correlación entre la recombinación y la diversidad (polimorfismo y divergencia)

Se ha observado una fuerte y consistente correlación entre el polimorfismo, medido como  $\pi$  ó  $\theta$  en los sitios 4-veces degenerados, y el entrecruzamiento, medido como  $c$  o  $\rho$  para cada brazo cromosómico (Tabla 3.8). Además, la correlación en autosomas es más intensa ( $r = 0,6105$ ) que en el cromosoma X ( $r = 0,4757$ ). La correlación entre la recombinación y  $\theta$  siempre es mayor que la observada con  $\pi$  (Tabla 3.8). En cuanto a las dos medidas de recombinación, la correlación es mayor con  $\rho$  en todos los brazos cromosómicos. Nuestro objetivo es evaluar si la correlación que existe entre estas dos variables desaparece a partir de un valor de recombinación dado. Asumimos que donde sea la correlación mayor, más difícil será observar falta de correlación en cualquier región genómica. Por tanto, usar las variables que muestran una correlación mayor es un criterio conservador. Por eso, en los análisis posteriores, como medida de polimorfismo usaremos  $\theta$ , pero también analizaremos los resultados tanto para  $c$  como para  $\rho$ , puesto que teóricamente la comparación entre ambos puede darnos información adicional si asumimos que  $\rho$  es la medida histórica de recombinación y  $c$  la medida actual de recombinación.

Siempre que se observa una correlación entre la recombinación y la divergencia, ésta es negativa (Tabla 3.8). *A priori*, no se espera que exista una correlación entre la divergencia y la recombinación a menos que la recombinación fuera mutagénica. En tal caso, esperaríamos una correlación positiva entre ambas medidas (Begun y Aquadro 1992) y no negativa. La reducida calidad de los alineamientos de las regiones repetitivas que abundan en las regiones centroméricas podrían generar una correlación negativa espuria como la que se observa. Por ello, se ha repetido el cálculo de la correlación quitando la región centromérica (un tercio de

cada brazo cromosómico). En estas nuevas estimas desaparece la correlación entre la recombinación y la divergencia (Tabla 3.8).

**Tabla 3.8.** Correlación de la recombinación con el polimorfismo y con la divergencia en cada brazo cromosómico y el total del genoma

Cromosoma	$\theta - c$		$\pi - c$		$\theta - \rho$		$\pi - \rho$	
	R	valor p	R	valor p	R	valor p	R	valor p
Total	0,5299	<2,2·10 <sup>-16</sup>	0,5066	<2,2·10 <sup>-16</sup>	0,6360	<2,2·10 <sup>-16</sup>	0,5811	<2,2·10 <sup>-16</sup>
2L	0,5498	<2,2·10 <sup>-16</sup>	0,4751	5,166·10 <sup>-14</sup>	0,7165	<2,2·10 <sup>-16</sup>	0,6292	<2,2·10 <sup>-16</sup>
2R	0,6674	<2,2·10 <sup>-16</sup>	0,6845	<2,2·10 <sup>-16</sup>	0,8317	<2,2·10 <sup>-16</sup>	0,7918	<2,2·10 <sup>-16</sup>
3L	0,6761	<2,2·10 <sup>-16</sup>	0,6405	<2,2·10 <sup>-16</sup>	0,7720	<2,2·10 <sup>-16</sup>	0,7282	<2,2·10 <sup>-16</sup>
3R	0,5598	<2,2·10 <sup>-16</sup>	0,5440	<2,2·10 <sup>-16</sup>	0,7442	<2,2·10 <sup>-16</sup>	0,6881	<2,2·10 <sup>-16</sup>
X	0,4757	<2,2·10 <sup>-16</sup>	0,3551	7,069·10 <sup>-8</sup>	0,4911	<2,2·10 <sup>-16</sup>	0,2924	1,247·10 <sup>-5</sup>
Autosomas	0,6105	<2,2·10 <sup>-16</sup>	0,5850	<2,2·10 <sup>-16</sup>	0,7718	<2,2·10 <sup>-16</sup>	0,7175	<2,2·10 <sup>-16</sup>

Cromosoma	$k - c$		$k - \rho$		$k - c$ sin centrómero		$k - \rho$ sin centrómero	
	R	valor p	R	valor p	R	valor p	R	valor p
Total	-0,0803	0,0059	-0,0660	0,0241	-	-	-	-
2L	-0,0972	0,1469	-0,1338	0,0459	-0,0915	0,2605	-0,0745	0,3602
2R	-0,2099	0,0022	-0,1020	0,1417	-0,0789	0,3504	0,1138	0,1806
3L	-0,2107	0,0010	-0,1389	0,0308	-0,1211	0,1224	-0,0011	0,9878
3R	-0,0310	0,6056	-0,0695	0,2493	0,0984	0,1813	0,1070	0,1460
X	-0,0181	0,7893	-0,0273	0,6882	-0,0865	0,2957	-0,1119	0,1771
Autosomas	-0,1109	0,0006	-0,0998	0,0021	-	-	-	-

Tabla superior: correlación de Spearman y valor  $p$  entre las dos estimas de recombinación ( $c$  y  $\rho$ ) y dos medidas de polimorfismo ( $\theta$  y  $\pi$ ) de los sitios 4 veces degenerados. Tabla inferior: correlación de Spearman y valor  $p$  entre las dos estimas de recombinación ( $c$  y  $\rho$ ) y la divergencia ( $k$ ) de los sitios 4 veces degenerados. Las correlaciones de ambas tablas se han estimado para cada brazo cromosómico en ventanas de 100kb. Las columnas denominadas sin centrómero son la correlación entre recombinación y divergencia tras excluir el tercio centromérico de cada brazo cromosómico.

### 3.3.3. Desacoplamiento entre polimorfismo y recombinación

Para detectar regiones donde la selección ligada es una fuerza evolutiva significativa hemos analizado la correlación entre recombinación y polimorfismo para cada cromosoma. Debido a que la divergencia mostró cierto efecto significativo sobre el polimorfismo en el análisis estadístico realizado (véase apartado 3.2), las estimas del polimorfismo de los sitios 4-veces degenerados se han corregido por la divergencia de los sitios 4-veces degenerados. Esta normalización suele emplearse para corregir el posible efecto de los cambios en la tasa de mutación (Lynch 2007). Para cada cromosoma, hay una correlación curvilínea entre estas dos variables. A partir de un valor de recombinación dado, la pendiente de la regresión lineal entre estas dos variables cambia notablemente. En un estudio previo (Mackay *et al.* 2012), se estimó a

partir de medidas de recombinación clásicas (Fiston-Lavier *et al.* 2010) que el valor umbral de recombinación podía estar en torno a 2 *cM/Mb*. En este trabajo, se ha utilizado un método de búsqueda iterativo para estimar el punto de truncamiento de recombinación para cada brazo cromosómico (véase apartado 2.4 de Materiales y Métodos). El valor de  $r^2$  indica que el modelo de regresión segmentada se ajusta mejor que una regresión lineal simple (Tabla 3.9). Debido a que en el modelo de regresión segmentada se añade una variable más, se probó si el modelo de regresión segmentada seguía siendo el más adecuado según el criterio de información de Akaike (AIC, Akaike 1974; apartado 2.7 de Materiales y Métodos). Aplicando AIC se confirma que el modelo de regresión segmentada explica mejor la relación entre el polimorfismo y la recombinación que una regresión lineal simple.

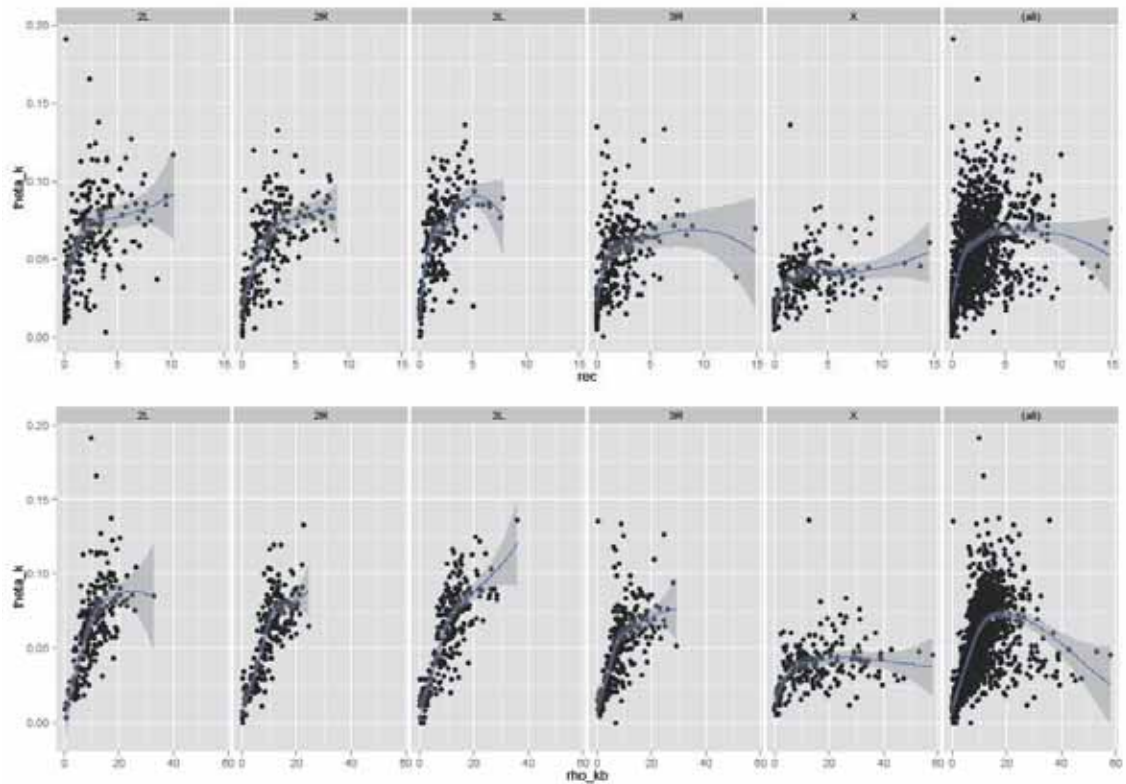
**Tabla 3.9.** Regresión lineal versus regresión segmentada entre el polimorfismo y la recombinación

<b>a) 100kb</b>						
	<b>Regresión lineal</b>			<b>Regresión segmentada</b>		
	<b><math>r^2</math> múltiple</b>	<b>valor <math>p</math></b>	<b>AIC</b>	<b><math>r^2</math> múltiple</b>	<b>valor <math>p</math></b>	<b>AIC</b>
2L	0,3888	$< 2,2 \cdot 10^{-16}$	-1054,65	0,4812	$< 2,2 \cdot 10^{-16}$	-1087,178
2R	0,5778	$< 2,2 \cdot 10^{-16}$	-1087,894	0,6724	$< 2,2 \cdot 10^{-16}$	-1136,897
3L	0,5892	$< 2,2 \cdot 10^{-16}$	-1212,587	0,6493	$< 2,2 \cdot 10^{-16}$	-1246,567
3R	0,3367	$< 2,2 \cdot 10^{-16}$	-1359,303	0,4209	$< 2,2 \cdot 10^{-16}$	-1392,752
X	0,0997	$2,315 \cdot 10^{-6}$	-1194,45	0,2273	$8,652 \cdot 10^{-12}$	-1223,308

<b>b) 10kb</b>						
	<b>Regresión lineal</b>			<b>Regresión segmentada</b>		
	<b><math>r^2</math> múltiple</b>	<b>valor <math>p</math></b>	<b>AIC</b>	<b><math>r^2</math> múltiple</b>	<b>valor <math>p</math></b>	<b>AIC</b>
2L	0,1237	$< 2,2 \cdot 10^{-16}$	-3261,091	0,1852	$< 2,2 \cdot 10^{-16}$	-3363,246
2R	0,1932	$< 2,2 \cdot 10^{-16}$	-3251,815	0,2206	$< 2,2 \cdot 10^{-16}$	-3299,613
3L	0,2121	$< 2,2 \cdot 10^{-16}$	-3089,775	0,2513	$< 2,2 \cdot 10^{-16}$	-3163,415
3R	0,1565	$< 2,2 \cdot 10^{-16}$	-4276,584	0,2026	$< 2,2 \cdot 10^{-16}$	-4378,750
X	0,0179	$7,204 \cdot 10^{-7}$	-3703,258	0,0657	$< 2,2 \cdot 10^{-16}$	-3770,372

Comparación del resultado de la regresión lineal y la regresión segmentada entre el polimorfismo de los sitios 4 veces degenerados y la recombinación en ventanas no solapantes de 100kb (a) y 10kb (b). Se dan la  $r^2$  múltiple, *valor p* y el valor AIC.



**Figura 3.9.** Regresión local suavizada entre el polimorfismo y la recombinación

Regresión local suavizada entre el polimorfismo corregido por la divergencia de los sitios 4-veces degenerados y las dos medidas de recombinación,  $c$  (gráfica superior) y  $\rho$  (gráfica inferior) en ventanas solapantes de 100kb.

El valor umbral de recombinación distingue entre una relación lineal bien definida (por debajo del valor umbral) y una relación no distinta de cero (por encima) entre el polimorfismo de los sitios 4-veces degenerados y la recombinación (Figura 3.9, Figura 3.10). Interpretamos la correlación positiva por debajo del valor umbral de recombinación como resultado de la selección ligada, ya sea por arrastre debido a la selección positiva o al arrastre por selección negativa, que abarca un tamaño de región más allá del tamaño de la ventana. Hemos denominado a estas regiones como **Bloques de Selección Ligada (LSB)**. Por el contrario, por encima del valor umbral el polimorfismo y la recombinación se comportan de forma independiente para esta resolución de ventana. Hemos denominado a estas regiones como **Bloques sin Selección Ligada (NLSB)** y asumimos que la selección ligada es una fuerza despreciable en estas regiones.



**Tabla 3.10.** *Valores estimados del umbral de recombinación*

a) <i>1Mb</i>	Valor umbral	Ordenada en el origen			Pendiente		
		Valor	IC en 2.50%	IC en 97.50%	Valor	IC en 2.50%	IC en 97.50%
2L	<25,98	0,0041	0,239	0,0081	$3,290 \cdot 10^{-6}$	3,442	0,7296
	>25,98	0,0878	3,404	0,0030	$-0,313 \cdot 10^{-6}$	-0,353	0,7283
2R	<19,73	0,0049	-0,0513	0,0610	$2,990 \cdot 10^{-6}$	$0,442 \cdot 10^{-6}$	$5,530 \cdot 10^{-6}$
	>19,73	0,0509	0,0237	0,0780	$0,960 \cdot 10^{-6}$	$-0,159 \cdot 10^{-6}$	$2,080 \cdot 10^{-6}$
3L	<27,31	0,0044	-0,2164	0,2253	$3,000 \cdot 10^{-6}$	$-4,150 \cdot 10^{-6}$	$10,100 \cdot 10^{-6}$
	>27,31	0,1580	0,0475	0,2680	$-2,342 \cdot 10^{-6}$	$-5,900 \cdot 10^{-6}$	$1,220 \cdot 10^{-6}$
3R	<16,34	0,0027	-0,0489	0,0543	$3,180 \cdot 10^{-6}$	$0,549 \cdot 10^{-6}$	$5,810 \cdot 10^{-6}$
	>16,34	0,0666	0,0421	0,0911	$-0,025 \cdot 10^{-6}$	$-1,120 \cdot 10^{-6}$	$1,070 \cdot 10^{-6}$
X	<14,55	0,0166	-0,0062	0,0393	$1,680 \cdot 10^{-6}$	$0,098 \cdot 10^{-6}$	$3,250 \cdot 10^{-6}$
	>14,55	0,0327	0,0235	0,0418	$0,247 \cdot 10^{-6}$	$-0,037 \cdot 10^{-6}$	$0,531 \cdot 10^{-6}$

b) <i>100kb</i>	Valor umbral	Ordenada en el origen			Pendiente		
		Valor	IC en 2.50%	IC en 97.50%	Valor	IC en 2.50%	IC en 97.50%
2L	<13,33	0,0297	0,0116	0,0479	0,0261	0,0135	0,0387
	>13,33	0,0649	0,0574	0,0723	0,0023	0,0005	0,0042
2R	<10,98	0,0123	-0,0327	0,0574	0,0054	0,0026	0,0081
	>10,98	0,0662	0,0443	0,0881	0,0009	-0,0003	0,0021
3L	<17,98	0,0140	-0,0626	0,0906	0,0046	0,0012	0,0081
	>17,98	0,0543	0,0161	0,0924	0,0016	$-2,27 \cdot 10^{-5}$	0,0033
3R	<11,46	0,0132	-0,0240	0,0504	0,0051	0,0026	0,0075
	>11,46	0,0510	0,0329	0,0690	0,0010	$-6,11 \cdot 10^{-5}$	0,0020
X	<5,17	0,0129	0,0012	0,0247	0,0052	0,0027	0,0077
	>5,17	0,0372	0,0334	0,0410	$1,62 \cdot 10^{-4}$	$-1,34 \cdot 10^{-5}$	0,0003

c) <i>10kb</i>	Valor umbral	Ordenada en el origen			Pendiente		
		Valor	IC en 2.50%	IC en 97.50%	Valor	IC en 2.50%	IC en 97.50%
2L	<26,9343	0,0377	-0,2703	0,3457	0,0037	0,0021	0,0053
	>26,9343	0,1418	-0,1652	0,4488	0,0001	-0,0005	0,0008
2R	<34,6558	0,0439	-0,0613	0,1491	0,0042	0,0019	0,0066
	>34,6558	0,1242	0,0720	0,1765	0,0011	$-3,86 \cdot 10^{-5}$	0,0022
3L	<33,9319	0,0370	-0,3189	0,3928	0,0051	0,0029	0,0073
	>33,9319	0,1471	-0,2057	0,4999	0,0010	$-2,75 \cdot 10^{-5}$	0,0021
3R	<25,3303	0,0274	-0,0220	0,0767	0,0049	0,0035	0,0063
	>25,3303	0,1233	0,0992	0,1473	0,0006	$-1,13 \cdot 10^{-5}$	0,0012

X	<8,5767	0,0241	-0,2133	0,2615	0,0051	0,0024	0,0078
	>8,5767	0,0841	-0,1528	0,3210	$-7,14 \cdot 10^{-5}$	$-2,94 \cdot 10^{-4}$	0,0001

Resultados de la regresión segmentada entre el polimorfismo de los sitios 4-veces degenerados y la recombinación en ventanas no solapantes de a) 1Mb, b) 100kb y c) 10kb. Valor umbral en  $\rho$  por kb. IC: Intervalo de confianza.

### 3.4. Mapa de LSB

Hemos localizado las regiones de cada brazo cromosómico que se sitúan por debajo y por encima del valor umbral de recombinación estimado (Figura 3.10, Tabla suplementaria 2). Aunque cada brazo cromosómico tiene su distribución propia de diversidad, en general es en los extremos de los brazos cromosómicos donde la recombinación está ausente o muy disminuida y es allí donde se concentran los grandes LSB. Por el contrario, en la región central de los brazos cromosómicos hay más regiones en los que la recombinación parece ser lo suficientemente elevada como para evitar los efectos de la selección ligada, abundando por tanto los NLSB.

#### 3.4.1. Patrón de LSB en autosomas y el cromosoma X

Considerando ventanas de 100kb y con  $\rho$  como medida de recombinación (Figura 3.10), la selección ligada afecta alrededor del 73% de los autosomas. En el cromosoma X, sin embargo, la selección ligada afecta sólo alrededor del 23% del brazo cromosómico (Tabla 3.11). Estas diferencias pueden deberse en parte a la mayor tasa de recombinación (1,36 veces) y del parámetro de recombinación poblacional ( $\rho$ ) (1,52 veces) que tiene el cromosoma X respecto a los autosomas (Tabla 3.7). Como la recombinación en *Drosophila melanogaster* sólo ocurre en hembras, al considerar la recombinación media poblacional debe corregirse por  $\frac{1}{2}$  y  $\frac{2}{3}$  en los autosomas y el cromosoma X, respectivamente. Además, teóricamente se considera que el censo efectivo del cromosoma X es  $\frac{3}{4}$  respecto a los autosomas. Teniendo en cuenta estas dos correcciones, en la comparación de  $\rho$  entre autosomas y X, las estimas de los autosomas habría que multiplicarlas por  $\frac{1}{2}$  y las del cromosoma X por  $\frac{1}{2}$  también ( $\frac{3}{4} \cdot \frac{2}{3}$ ). Es decir, que para  $\rho$  no esperamos diferencias teóricas entre los autosomas y el X. Sin embargo, observamos un valor de  $\rho$  1,52 veces mayor en el cromosoma X. Esto podría ser debido a un incremento en la tasa de recombinación o a un incremento en el censo efectivo del X (los dos parámetros de los que depende  $\rho$ ,  $\rho = 4N_e r$ ). La tasa de recombinación ( $c$ ), sin embargo, se debe normalizar multiplicando por  $\frac{1}{2}$  y  $\frac{2}{3}$  en los autosomas y el cromosoma X, respectivamente. Al hacerlo, las diferencias son todavía más notables (*valor p* de la prueba de Wilcoxon = 0) siendo  $c_X$  1,81 veces  $c_A$ , es decir, un incremento más allá de los  $\frac{4}{3}$  teóricos esperados ( $r_X = \frac{4}{3} r_A$ ). Parece, por tanto, que el incremento de 1,52 en la estima de  $\rho$  del cromosoma X se debe al incremento de la recombinación en este cromosoma. A pesar de que las estimas del parámetro de recombinación y de la tasa de recombinación son dos medidas completamente independientes en cuanto al método de estimación, dan sin embargo unos resultados consistentes que muestran que la

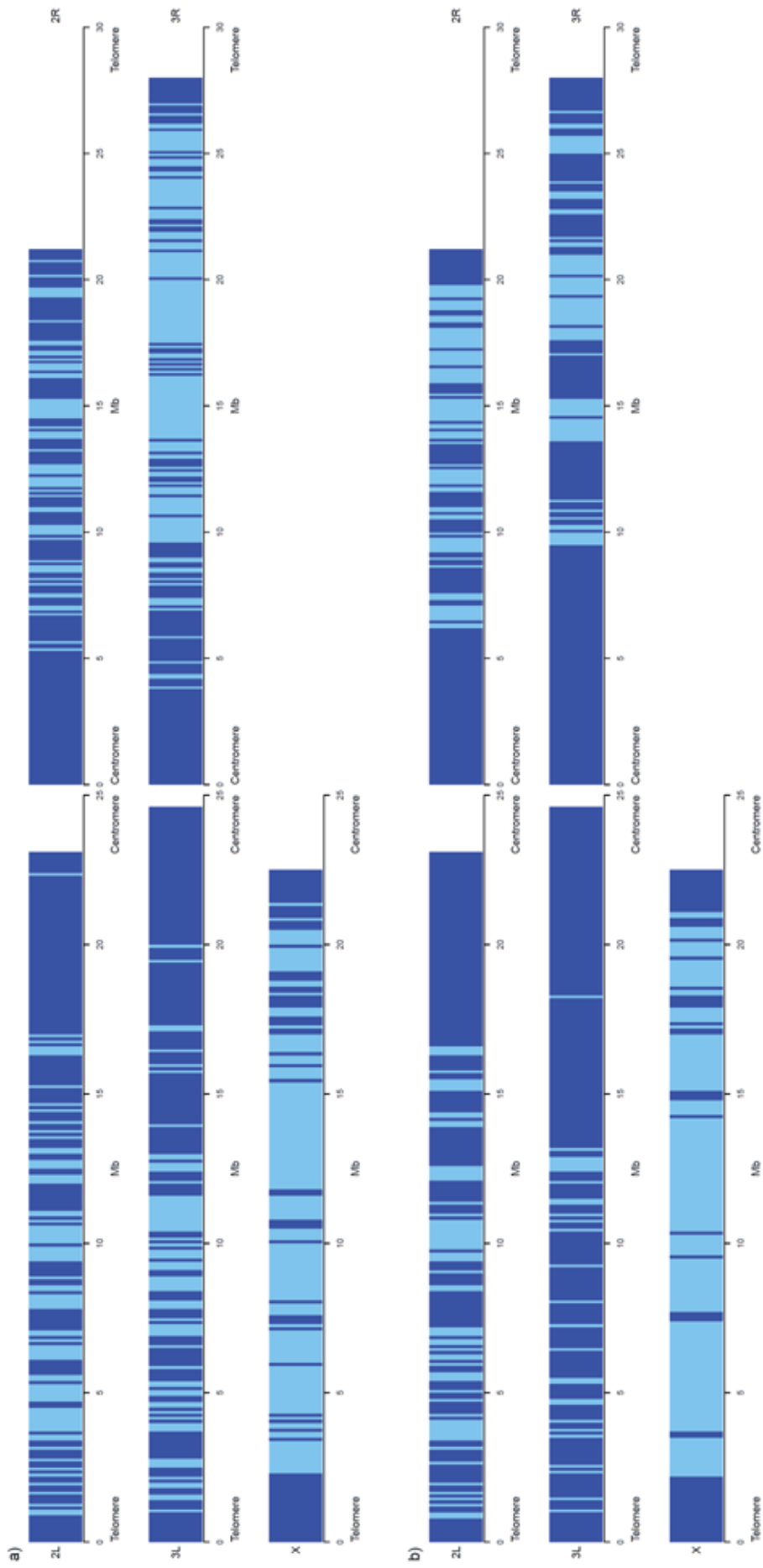
recombinación del cromosoma X es claramente superior a la de los autosomas. Por tanto, con los resultados obtenidos hasta ahora, podemos suponer que la mayor tasa de recombinación observada en el X es responsable de la abundancia de regiones donde la selección ligada está ausente o puede considerarse despreciable (NLSB) en las ventanas de 100kb analizadas.

**Tabla 3.11.** *Proporción de sitios de los LSB y los NLSB*

Brazo cromosómico	100kb con $c$			100kb con $\rho$		
	Valor umbral	% LSB	% NLSB	Valor umbral	% LSB	% NLSB
2L	2,81	64,50	35,50	13,33	70,18	29,82
2R	3,36	69,34	30,66	10,98	61,35	38,65
3L	2,42	71,14	28,86	17,98	88,98	11,02
3R	1,31	46,07	53,93	11,46	72,36	27,64
X	1,54	33,33	66,67	5,17	22,54	77,46

Valor umbral, % del cromosoma formado por LSB y % del cromosoma formado por NLSB estimado a partir del mapa basado en la tasa de recombinación ( $c$ ) y en el parámetro de recombinación ( $\rho$ ).

Además de la longitud total que ocupan los LSB y lo NLSB en cada cromosoma, otra de las características que permite analizar el mapa elaborado es la longitud promedio de los distintos bloques. El promedio del tamaño de los LSB en los autosomas es de 764kb, y fluctúa desde 100kb (el mínimo alcanzable dado que la ventana de observación es de 100kb) a 9,5Mb, el tercio centromérico del brazo cromosómico del 3R (Tabla 3.12). Los NLSB son en promedio 2,8 veces menores (~247kb) que los LSB y pueden abarcar desde 100kb a 1,1Mb. Si no se consideran los enormes bloques centroméricos y teloméricos (Tabla suplementaria 3), la longitud promedio de los LSB sigue siendo mayor que la de los NLSB, 452kb vs 247kb respectivamente, pero las diferencias son menores, reduciéndose de 2,8 veces a 1,65 veces la longitud promedio de los NLSB. Sin embargo, como ya no debería extrañar a estas alturas, el cromosoma X exhibe el patrón opuesto: los NLSB son en promedio 2,9 veces mayores que los LSB. El tamaño promedio de los NLSB es de ~1,2Mb, pudiendo abarcar desde 100kb a 3,8Mb, mientras que el tamaño promedio de los LSB es de ~400kb: de 100kb a 2,2Mb. Cuando en el cromosoma X se excluyen las regiones centroméricas y teloméricas, las diferencias aumentan, puesto que la longitud promedio de los LSB se reduce a ~185kb (~6,4 veces menos que la longitud de los NLSB). Resumiendo: (i) las regiones centroméricas y teloméricas están formadas por grandes LSB, (ii) las regiones centrales de los autosomas están compuestas principalmente de LSB que en promedio son 1,65 veces más largos que los NLSB, pero (iii) la región central del X está esencialmente formada por NLSB que son en promedio 6,4 veces mayores que los LSB. Estas diferencias podrían explicarse en gran parte por la fracción mayor de recombinación del cromosoma X respecto a lo autosomas.



**Figura 3.10.** Mapa de bloques de selección ligada (LSB) del genoma de *D. melanogaster* cartografiado a partir de ventanas de 100kb

Mapa de LSB basado en a) la tasa de recombinación  $\rho$  y b) el parámetro poblacional de recombinación  $\rho$ , para cada brazo cromosómico. LSB = azul claro y NLSB = azul oscuro.

**Tabla 3.12.** Número y tamaño de los LSB

Cromosoma	kb	Número		Tamaño mínimo (kb)		Tamaño máximo (kb)		Tamaño medio (kb)		
		LSB	NLSB	LSB	NLSB	LSB	NLSB	LSB	LSB-CT	NLSB
2L	10	228	227	10	10	1.070	210	68,77	63,36	29,91
	100	30	29	100	100	6.500	1.000	553,33	332,14	224,14
	1.000	4	3	1.000	1.000	12.000	2.000	3.600	1.667	1250
2R	10	161	160	10	10	1.400	80	112,17	104,09	19,25
	100	24	23	100	100	6.200	900	545,83	250	352,17
	1.000	3	2	1.000	2.000	6.000	8.000	2250	1.000	4.000
3L	10	207	206	10	10	4.190	260	98,45	96,93	20,19
	100	22	21	100	100	6.300	500	990,91	725	133,33
	1.000	3	2	2.000	1.000	11.000	1.000	7.333	9.000	1.000
3R	10	234	233	10	10	3360	290	92,91	87,5	26,44
	100	21	20	100	100	9500	1.100	966,67	500	385
	1.000	5	4	1.000	2.000	10.000	7.000	3.750	1.667	3.250
X	10	269	268	10	10	400	420	26,43	25,24	57,13
	100	15	14	100	100	2.200	3.800	400	184,62	1.178,57
	1.000	2	1	1.000	19.000	2.000	19.000	1.500	0	19.000

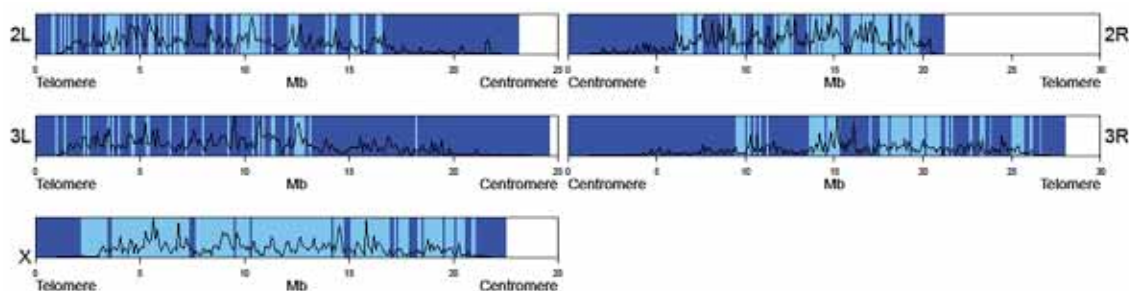
Número, tamaño mínimo, tamaño máximo y tamaño medio de los LSB y NLSB de los mapas estimados a 10kb, 100kb y 1Mb.

### 3.4.2. Fiabilidad del mapa de LSB

Hemos utilizado dos medidas de recombinación para realizar nuestro análisis,  $\rho$  y  $c$ . Cada medida representa de modo distinto el tiempo y a los individuos de una misma población (véase sección 4.7). Por lo tanto, además del mapa basado en el parámetro de recombinación ( $\rho$ ), se construyó un mapa alternativo de LSB usando la tasa de recombinación ( $c$ ). Comparar los resultados entre un mapa y otro nos puede dar información adicional sobre la dinámica de la tasa de recombinación. Ya sea porque  $\theta$  y  $\rho$  son ambas medidas históricas afectadas por el censo efectivo de la población o porque la recombinación actual tiene un coeficiente de variación mayor (Tabla 3.7) que implica más ruido que la recombinación histórica, la correlación entre las dos variables  $\theta$  y  $\rho$  es más fuerte que la correlación entre  $\theta$  y  $c$  (Tabla 3.8). Estas diferentes correlaciones harán que se observe un menor umbral de recombinación y una menor proporción de sitios formando LSB cuando se utiliza  $c$  como medida de recombinación en vez de  $\rho$ . En cualquier caso, este efecto debería ser similar en todos los cromosomas o mayor en aquellos brazos cromosómicos con mayor coeficiente de variación. Si este (u otros) sesgo no es muy pronunciado, la similitud entre ambos mapas nos indicaría que la recombinación y la selección ligada no han variado drásticamente a lo largo del tiempo. Por el contrario, si se hubieran

observado diferencias sustanciales en ambos mapas nos llevaría a suponer que la recombinación ha cambiado desde la separación de ambas especies analizadas.

Observamos que aunque hay regiones específicas que difieren entre los mapas  $c$  y  $\rho$ , el patrón general es consistente entre las dos medidas de recombinación (Figura 3.10). Los bloques del mapa basado en  $\rho$  están estrechamente correlacionados con las estimas de  $c$  en ventanas de 100kb (R de *Spearman* = 0,54, valor  $p$  = 0, Figura 3.11). Es decir,  $c$  parece predecir de forma adecuada los LSB y NLSB calculados a partir de  $\rho$ . Una vez más, el patrón de los autosomas difiere respecto al del cromosoma X. En el mapa basado en  $c$ , hay un ~10% menos de LSB que en el mapa basado en  $\rho$  (LSB basado en  $c$ : ~63%, LSB basadas en  $\rho$ : ~73%, Tabla 3.11). En el cromosoma X, sin embargo, los LSB obtenidos a partir de  $c$  abarcan un 10% más del cromosoma que en el mapa basado en  $\rho$  (LSB basado en  $c$ : ~33%, LSB basadas en  $\rho$ : ~23%). Globalmente, con el parámetro de recombinación,  $\rho$ , los LSB ocupan un 5% más del genoma comparado en el mapa basado en  $c$  (~63% *versus* ~57%, respectivamente).



**Figura 3.11.** Tasa de recombinación a lo largo del cromosoma superpuesta el mapa de LSB

Mapa de LSB basado en el parámetro de recombinación ( $\rho$ ); LSB en azul claro y BLNS en azul oscuro. La línea negra es la tasa de recombinación ( $c$ ) en ventanas no solapantes de 100kb a lo largo de los brazos cromosómicos.

Si analizamos con más detalle el mapa de LSB, observamos que las mayores diferencias entre uno y otro mapa se dan en el cromosoma 3 (Figura 3.10, Tabla 3.11). En este caso, los LSB representan una proporción mayor del cromosoma (3L: incremento de 17,84 y 3R: incremento de 26,29) cuando se utiliza  $\rho$  como medida de recombinación (Tabla 3.11). El brazo cromosómico 3L es en ambos mapas el brazo con mayor proporción de LSB, es decir, el brazo cromosómico con mayor efecto de la selección ligada. Sin embargo, el 3R pasa de ser uno de los cromosomas con mayor efecto de la selección ligada, según  $\rho$ , al autosoma con menor efecto según  $c$ . Sabemos que el coeficiente de variación del 3R con  $c$  es ~1 (Tabla 3.7). Esta gran varianza de la tasa de recombinación incrementa el error asociado a la estimación de  $c$  en este cromosoma lo que podría reducir la fuerza de la correlación entre el polimorfismo y la recombinación haciendo disminuir los LSB observados. Este incremento en el coeficiente de variación del 3R podría estar causado por el mayor número de inversiones que alberga este cromosoma respecto al resto de los cromosomas de *D. melanogaster* (Corbett-Detig y Hartl

2012), y podría explicar todo o parte del patrón discordante entre una y otra medida. Puesto que las inversiones alteran la tasa de recombinación de las regiones afectadas (Navarro *et al.* 2000), su mayor abundancia en el 3R podría explicar los cambios recientes en la tasa de recombinación de este brazo cromosómico. El resto de las diferencias entre cromosomas son mucho más pequeñas (2L: incremento de 5,68, 2R: reducción de 7,99 y X: reducción de 10,79; Tabla 3.11) lo que nos estaría indicando que no ha habido cambios sustanciales en la tasa de recombinación a lo largo del tiempo.

### 3.4.3. Densidad génica y el mapa de LSB

La proporción de sitios que forman los LSB representan ~63% del genoma y contienen un ~65% de los genes en el mapa basado en  $\rho$ . Para el mapa basado en  $c$ , los LSB suponen un ~56,9% y contienen ~56% de los genes (Tabla 3.13). Estas diferencias no son significativas de acuerdo con el análisis de remuestreo *bootstrap* realizado (Tabla 3.14). Por lo tanto, los NLSB no están enriquecidos en genes y, consecuentemente, la densidad génica no parece estar asociada a los LSB.

**Tabla 3.13.** Número y proporción de genes en los LSB estimados

<i>a) 1Mb</i>								
Cromosoma	Número de genes			% de genes		% de pb		% genes - % pb
	Total	LSB	NLSB	LSB	NLSB	LSB	NLSB	
2L	2.755	1.989	766	72,20	27,80	78,26	21,74	-6,06
2R	3.079	1.144	1.935	37,15	62,85	42,86	57,14	-5,71
3L	2.836	2.632	204	92,81	7,19	91,67	8,33	1,14
3R	3.536	1.799	1.737	50,88	49,12	51,85	48,15	-0,97
X	2.288	298	1.990	13,02	86,98	13,64	86,36	-0,62

<i>b) 100kb con <math>\rho</math></i>								
Cromosoma	Número de genes			% de genes		% de pb		% genes - % pb
	Total	LSB	NLSB	LSB	NLSB	LSB	NLSB	
2L	2.755	2.008	747	72,86	27,11	70,18	29,82	2,68
2R	3.079	1.984	1.095	64,44	35,56	61,35	38,65	3,09
3L	2.836	2.609	227	92,00	8,00	88,98	11,02	3,02
3R	3.536	2.587	949	73,16	26,84	72,36	27,64	0,80
X	2.288	526	1.762	22,99	77,01	22,54	77,46	0,46

c) 100kb con $c$								
Cromosoma	Número de genes			% de genes		% de pb		% genes - % pb
	Total	LSB	NLSB	LSB	NLSB	LSB	NLSB	
2L	2.755	1.703	1.052	61,81	38,19	64,50	35,50	-2,69
2R	3.079	2.145	934	69,67	30,33	69,34	30,66	0,33
3L	2.836	2.046	790	72,14	27,86	71,14	28,86	1,00
3R	3.536	1.603	1.933	45,33	54,67	46,07	53,93	-0,74
X	2.288	713	1.575	31,16	68,84	33,33	66,67	-2,17

d) 10kb								
Cromosoma	Número de genes			% de genes		% de pb		% genes - % pb
	Total	LSB	NLSB	LSB	NLSB	LSB	NLSB	
2L	2.755	2.123	632	77,06	22,94	69,60	30,40	7,46
2R	3.079	2.760	319	89,64	10,36	86,45	13,55	3,19
3L	2.836	2.532	304	89,28	10,72	85,02	14,98	4,26
3R	3.536	2.996	540	84,73	15,27	80,20	19,80	4,53
X	2.288	754	1.534	32,95	67,04	26,67	73,33	6,28

Número y proporción de genes en los LSB y los NLSB basado en el mapa de LSB de 1Mb (a), de 100kb con  $\rho$  (b), de 100kb con  $c$  (c) y de 10kb. También se muestra la proporción de pb de los LSB y NLSB para facilitar la comparación.

**Tabla 3.14.** Diferencias entre el número de genes de los LSB y NLSB

Cromosoma	LSB		NLSB	
	2.5%	97.5%	2.5%	97.5%
2L	9	12	8	14
2R	12	16	11	15
3L	10	13	4	10,5
3R	11	14	11	14
X	6	10	9	11

Resultado del remuestreo *bootstrap* para el número de genes en los LSB y NLSB. Si los resultados solapan se considera que no hay diferencias significativas.

## 3.5. Características de los LSB

### 3.5.1. Razón $c/\mu$

Si suponemos que  $\theta$  es una estima de  $N_e\mu$  y  $\rho$  de  $N_e c$ , entonces el cociente entre  $N_e\mu / N_e c$  es una estima de  $c/\mu$ , la razón de la tasa de recombinación y la tasa de mutación por sitio nucleotídico. Un valor  $c/\mu$  bajo implica una mayor vulnerabilidad a los efectos de la selección ligada debido a la incapacidad de la recombinación para separar (independizar estadísticamente) dos o más nuevas mutaciones surgidas en una misma región genómica. Por el contrario, valores elevados de  $c/\mu$  pueden aliviar en gran medida los efectos producidos por la interferencia de



Hill-Robertson (iHR, Hill y Robertson 1968, Lynch 2007). Dividiendo las estimas de  $\rho$  y  $\theta$  de las posiciones 4 veces degeneradas, podemos obtener estimas de  $c/\mu$  de nuestros genomas y comprobar si esta razón difiere entre los LSB y los NLSB.

### 3.5.1.1. Estimaciones de la razón $c/\mu$ a escala genómica

Por primera vez se han estimado los valores de  $c/\mu$  a escala genómica en *Drosophila melanogaster*. La media genómica de la razón  $c/\mu$  es 1,1070, mucho menor a las estimas obtenidas previamente en uno o varios genes de esta especie (media de  $c/\mu = 3,8$ ; Hey y Warkeley 1997, Machado *et al.* 2002) (Tabla 3.15). Nuestra estima parece indicar que la fuerza relativa promedio de la mutación y la recombinación son similares.

**Tabla 3.15.** Descripción estadística de la razón  $c/\mu$

a)							
$c/\mu$	Genoma	Autosomas	2L	2R	3L	3R	X
Media ( $\bar{x}$ )	1,1070	0,9072	0,8582	0,8591	0,9311	0,9627	1,9870
Mediana ( $P_{50}$ )	0,8778	0,8227	0,7930	0,8334	0,8782	0,8207	1,5300

b) $\rho$								
Cromosoma	LSB				NLSB			
	$\bar{x}$	$P_{50}$	$P_{2,5}$	$P_{97,5}$	$\bar{x}$	$P_{50}$	$P_{2,5}$	$P_{97,5}$
2L	0,7508	0,7115	0,6521	0,7811	1,1190	1,0110	0,8849	1,0215
2R	0,7461	0,7204	0,6368	0,7707	1,0360	1,0570	0,8851	1,0872
3L	0,8808	0,8105	0,7092	0,8082	1,3100	1,2540	1,0570	1,2138
3R	0,8179	0,7311	0,7007	0,7507	1,3350	1,3070	1,0600	1,3000
X	0,4805	0,4275	0,3538	0,4681	2,4320	1,9600	1,7081	2,1274

c) $c$								
Cromosoma	LSB				NLSB			
	$\bar{x}$	$P_{50}$	$P_{2,5}$	$P_{97,5}$	$\bar{x}$	$P_{50}$	$P_{2,5}$	$P_{97,5}$
2L	0,7858	0,7332	0,6729	0,7692	0,9826	0,8840	0,7985	0,8849
2R	0,8151	0,7929	0,7175	0,8567	0,9579	0,9291	0,8233	0,9668
3L	0,8514	0,8095	0,7243	0,8882	1,1200	0,9851	0,9345	1,0759
3R	0,8123	0,7194	0,7023	0,7658	1,0880	1,0030	1,0103	1,2111
X	1,0630	0,7387	0,4187	0,7404	2,4240	1,9750	1,7635	2,2084

(a) Media y mediana de la razón  $c/\mu$  para todo el genoma, autosomas, y cada brazo cromosómico. Media, mediana y cuartiles de la razón  $c/\mu$  en los LSB y NLSB del mapa de 100kb basado en  $\rho$  (b) y  $c$  (c) por brazo cromosómico.

El valor medio estimado puede resultar engañoso, porque hay diferencias muy claras y significativas en los valores de  $c/\mu$  entre los LSB y NLSB para cada brazo cromosómico (excepto para el brazo cromosómico 2R con el mapa basado en  $c$ ; Tabla 3.15). En promedio, el

---

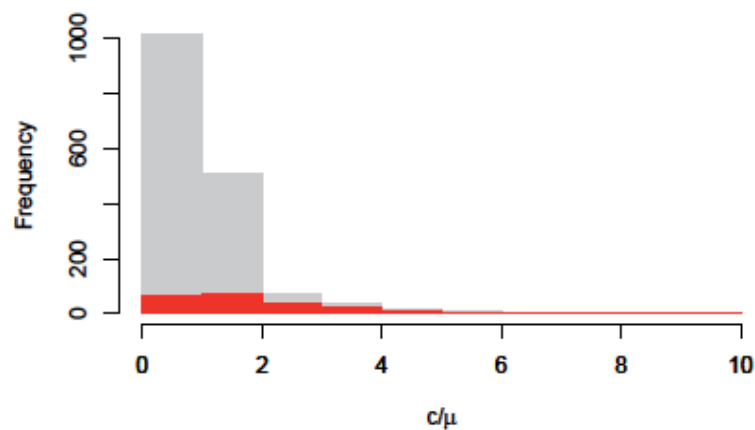
$c/\mu$  de los LSB es 0,7530 y 0,8187 para los mapas basados en  $\rho$  y  $c$ , respectivamente. Es decir, valores inferiores a 1 indican que la mutación predomina sobre la recombinación en estas regiones del genoma. Sin embargo, en los NLSB, la media es  $c/\mu = 1,5270$  y  $c/\mu = 1,4020$  para los mapas basados en  $\rho$  y  $c$ , respectivamente. El contraste no puede ser más claro entre ambos bloques, en los NLSB la fuerza de la recombinación es netamente superior a la fuerza de la mutación.

Cabe destacar las diferencias tan marcadas que existen, una vez más, entre el cromosoma X y el resto de los brazos cromosómicos. En promedio, el X tiene un  $c/\mu$  2,19 veces superior a los autosomas ( $c/\mu_x=1,9870$ ), por tanto la fuerza de la recombinación parece ser muy superior a la de la mutación en el cromosoma sexual (Tabla 3.15). En los NLSB, donde la recombinación predomina sobre la mutación (Tabla 3.15), el valor de  $c/\mu$  en el cromosoma X es también  $\sim 2$  veces superior a los autosomas (2,03 y 2,38 veces superior según el mapa basado en  $\rho$  y  $c$ , respectivamente). Pero en los LSB, mientras que para el mapa basado en  $\rho$  el cromosoma X muestra la  $c/\mu$  más baja de todos los brazos cromosómicos (X  $c/\mu_{LSB}$ : 0,4805, es decir, 1,66 veces menor que los autosomas), para el mapa basado en  $c$  continua siendo el más alto (X  $c/\mu_{LSB}$ : 1,063, o sea, 1,31 veces mayor que los autosomas) y el único en el que la mutación y la recombinación parecen tener un peso similar. La elaboración del mapa basado en  $\rho$  y las estimas obtenidas de  $c/\mu$  utilizan las mismas variables:  $\theta$  y  $\rho$ . Por tanto, es de esperar que el mapa obtenido con el parámetro poblacional de recombinación permita estimar de forma más precisa las regiones más o menos afectadas por la selección ligada si lo validamos con las estimas de  $c/\mu$ . Sin embargo, el mapa basado en  $c$  es independiente de  $\theta$  y  $\rho$ , y a pesar de mostrar ciertas diferencias, el patrón general es el mismo. Esta similitud parece indicar que nuestro método es efectivo separando las regiones afectadas y no afectadas por la selección ligada en ventanas de 100kb. Además, la alta recombinación que tiene el X se ve reflejada en estas medidas, poniendo de manifiesto que el efecto de la recombinación respecto a la mutación es aproximadamente el doble que en los autosomas.

### **3.5.1.2. Simulaciones del efecto de la razón $c/\mu$ sobre la selección ligada**

Quisimos analizar si los valores promedio de  $c/\mu$  observados en los NLSB son suficientemente elevados como para considerar que los efectos de la selección ligada son despreciables en estas regiones. Para ello simulamos el valor de  $\theta$  y de  $\alpha$  para distintos valores de  $c/\mu$  utilizando el software (SFS\_CODE) (Hernandez 2008) (véase Materiales y Métodos sección 2.11). Debido al alto coste computacional que requiere la simulación, no se pudieron estimar los valores de  $\theta$  y de  $\alpha$  bajo el supuesto de estricta independencia. Con el software usado, adecuado para simular sitios ligados, las estimas obtenidas bajo condiciones de estricta independencia utilizan unas aproximaciones matemáticas distintas que no son estrictamente comparables a las obtenidas bajo condiciones de ligamiento. Para obtener estimas de

independencia comparables a las estimas obtenidas bajo condiciones de ligamiento hay que ir incrementando los valores de recombinación hasta que las variables simuladas no varíen, lo que supone un gran coste computacional. El valor máximo que se simuló fue  $c/\mu = 10$ , un valor  $\sim 10$  veces superior al de la media genómica observada,  $\sim 2,5$  veces superior al máximo de los autosomas e igual al máximo observado en el cromosoma X (en ventanas de 100kb) (Figura 3.12). Pensamos que con  $c/\mu = 10$  obtenemos estimas muy similares a las que se obtendrían bajo condiciones de independencia (Figura 3.13), lo que nos permitiría contestar de forma adecuada a la pregunta planteada. La razón  $c/\mu$  se estimó para el entrecruzamiento y para la conversión génica de forma independiente, de modo que podemos evaluar el impacto que supone cada uno de las fuerzas de recombinación. Los parámetros de mutación, recombinación y selección utilizados para la simulación intentan ser los más próximos a los valores observados en *Drosophila* (Materiales y métodos sección 2.11; Schneider *et al.* 2011; Comeron *et al.* 2012).



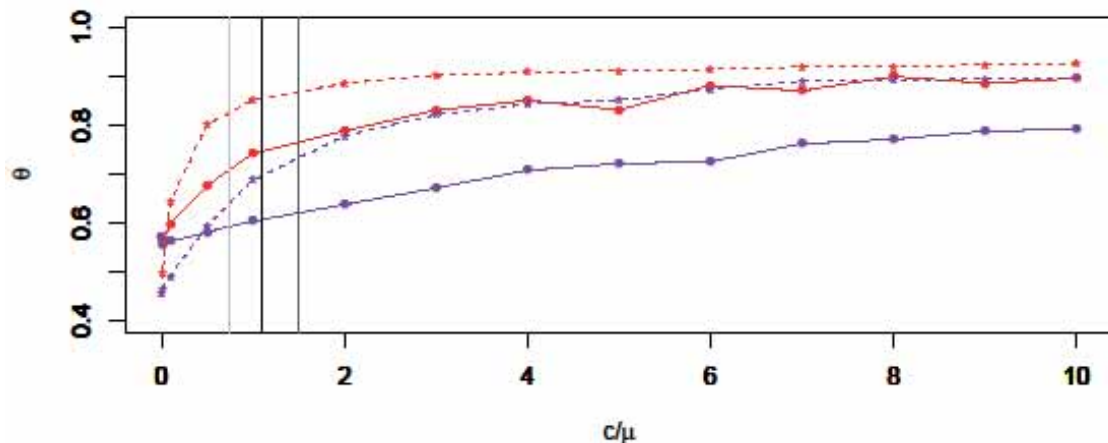
**Figura 3.12.** Distribución de la razón  $c/\mu$  observada

Distribución de los valores observados de  $c/\mu$  en ventanas de 100kb para todo el genoma (gris). La parte proporcional del X se destaca en rojo. A partir de  $c/\mu=4$  todos los valores son del cromosoma X.

Las simulaciones para distintos niveles de la tasa de recombinación muestran una curva asintótica en la que cada aumento de la razón  $c/\mu$  reduce el incremento relativo de la eficiencia de selección (si todas las demás variables son iguales,  $\alpha$  es un buen estimador de la eficiencia de selección) y la diversidad nucleotídica ( $\theta$ ) (Figura 3.13, Tabla 3.16). Según estos resultados, la selección ligada reduce  $\alpha$  de forma dramática para razones  $c/\mu$  menores a  $\sim 1$ . Para  $c/\mu = 1$ , hay un incremento del 39% del valor de  $\alpha$  respecto a los valores de entrecruzamiento igual a 0. Sin embargo, para valores de  $c/\mu = 10$ ,  $\alpha$  incrementa tan sólo un 8% respecto a los valores de  $c/\mu = 1$ . Es decir, que para valores de  $c/\mu > 1$ , el efecto de la selección ligada sobre la eficiencia de selección parece ser menor. Cuando  $c/\mu = 1$ ,  $\theta$  es un 17% mayor respecto a su valor cuando  $c/\mu = 0$  y este, a su vez, es un 16% menor respecto al máximo estimado ( $c/\mu = 10$ ). La selección ligada parece reducir los niveles de diversidad de forma más gradual que los valores que miden la eficiencia de selección, aunque la curva sigue siendo asintótica con una importante reducción

para valores  $c/\mu < 1$  (Figura 3.13, Tabla 3.16). Parece, por tanto, que la selección ligada afecta de forma más intensa los niveles de polimorfismo que la eficiencia de la selección. Esto sugiere que hay regiones donde la selección ligada puede dejar una huella sobre el polimorfismo al mismo tiempo que la eficiencia de la selección es alta o similar a la de una región donde la selección ligada es despreciable.

Con la simulación del entrecruzamiento, para las media de  $c/\mu$  observadas en los LSB (Tabla 3.15), representada con la línea gris claro en la Figura 3.13, los valores de  $\alpha$  están a tan sólo a 9% del máximo mientras que los de  $\theta$  están un 19% por debajo del máximo observado ( $c/\mu = 10$ ). Para los NLSB (la línea gris oscuro en la Figura 3.13),  $\alpha$  está a un 6% del máximo simulado y  $\theta$  está todavía un 13% por debajo del máximo observado. Las diferencias entre los LSB y NLSB son el doble de grandes para  $\theta$  (6%) que para  $\alpha$  (3%). Según la simulación, los valores observados de  $c/\mu$  en los NLSB no parecen ser suficientes para eliminar completamente los efectos deletéreos de la selección ligada. Sin embargo, la eficiencia de la selección no parece ser muy distinta entre los LSB y NLSB. Este resultado es simplemente indicativo, pues los valores simulados, aunque intentan reflejar lo descrito en poblaciones naturales, puede que no se ajusten a los valores reales del genoma que estamos analizando.



**Figura 3.13.** Simulación de los valores de  $\alpha$  y  $\theta$  en función de la razón  $c/\mu$

Estimas de  $\alpha$  (líneas discontinuas) y  $\theta * 1000$  (líneas sólidas) para distintos valores de  $c/\mu$  simulados. Además, la recombinación ( $c$ ) puede ser causada por entrecruzamiento (rojo) o conversión génica (morado). Las líneas verticales representan el promedio de  $c/\mu$ : (i) del genoma (negro), (ii) de los LSB (gris claro) y (iii) de los NLSB (gris oscuro).

Todo lo anterior es cierto cuando la fuerza principal de recombinación es el entrecruzamiento. Si la fuerza principal de recombinación simulada es la conversión génica, para cada valor de  $c/\mu$  las estimas de  $\alpha$  y  $\theta$  están en promedio un  $\sim 9,7\%$  y un  $\sim 13,5\%$  reducidas, respectivamente (Figura 3.13, Tabla 3.16). Es decir, el efecto de la selección ligada es mayor cuando la fuerza principal de recombinación es la conversión génica. Incluso a valores tan pequeños como  $c/\mu = 0,01$ , el entrecruzamiento parece que independiza los sitios de forma más

eficiente que la conversión génica, incrementando en un 6% y un 2% los valores de  $\alpha$  y  $\theta$ , respectivamente. Para que el valor de  $\alpha$  incremente un 39%, valor que se obtiene de  $c/\mu = 0$  a  $c/\mu = 1$  con el entrecruzamiento, con la conversión génica como fuerza de recombinación, el valor de  $c/\mu$  debería ser igual a 4. Para  $\theta$ , el incremento del 17% observado de  $c/\mu = 0$  a  $c/\mu = 1$  con el entrecruzamiento es obtenido para valores de  $c/\mu > 6$  cuando la conversión génica es la fuerza de recombinación. En resumen, como se podría suponer, aunque la conversión génica reduce también los efectos negativos de la selección ligada, es el entrecruzamiento la fuerza de recombinación que con mayor eficiencia (un 11,6% en promedio) independiza los sitios en el genoma.

**Tabla 3.16.** Simulación de los valores de  $\alpha$  y  $\theta$  en función de la razón  $c/\mu$

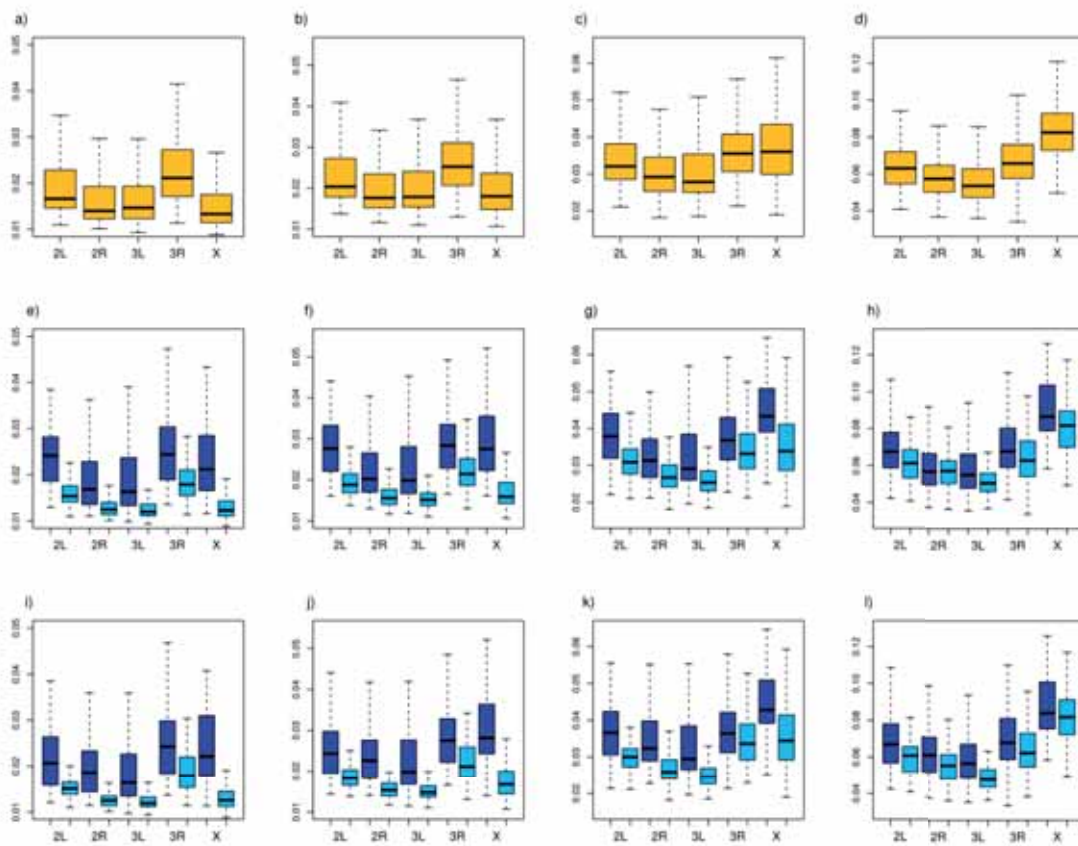
$c/\mu$	Entrecruzamiento		Conversión génica	
	$\alpha$	$\theta$	$\alpha$	$\theta$
0	0,4578	0,5717	0,4578	0,5717
0,01	0,4968	0,5668	0,4657	0,5546
0,1	0,6435	0,5978	0,4923	0,5631
0,5	0,8034	0,6764	0,5964	0,5831
1	0,8520	0,7425	0,6890	0,6045
2	0,8871	0,7891	0,7798	0,6386
3	0,9030	0,8316	0,8229	0,6723
4	0,9093	0,8518	0,8442	0,7089
5	0,9132	0,8324	0,8526	0,7227
6	0,9158	0,8822	0,8745	0,7264
7	0,9208	0,8719	0,8916	0,7632
8	0,9211	0,9007	0,8930	0,7723
9	0,9255	0,8875	0,8962	0,7890
10	0,9277	0,8969	0,8980	0,7933

Resultados de la simulación de  $\alpha$  y  $\theta$  (multiplicado por 1000) para distintos valores de  $c/\mu$ . La recombinación ha sido simulada como entrecruzamiento o conversión génica.

### 3.5.2. Patrón de desequilibrio de ligamiento ( $r^2$ )

El desequilibrio de ligamiento es la asociación no aleatoria de alelos en diferente *loci* y está inversamente correlacionado con la tasa de recombinación (Nachman 2002). Tanto  $c$  (*Spearman R*: -0,62; valor  $p = 0$ ) como  $\rho$  (*Spearman R*: -0,60; valor  $p = 0$ ) están fuerte y negativamente correlacionados con  $r^2$  cuando se estima en ventanas de 100kb. De acuerdo con los intervalos de confianza obtenidos mediante un remuestreo *bootstrap* (Tabla 3.17), los LSB, donde en promedio la recombinación es menor, presentan valores de  $r^2$  mayores que los NLSB (Figura 3.14), como se esperaba.

A diferencia de estudios previos (Wall *et al.* 2002; Mackay *et al.* 2012; Langley *et al.* 2012), la  $r^2$  promedio del cromosoma X es la menor de los promedios cromosómicos para tamaños de ventana de 100kb. Esta observación nos llevó a explorar más detalladamente nuestros valores de  $r^2$  y extendimos su estima a distintos tamaños de ventana (50kb, 10kb y 1kb). La posición relativa del cromosoma X varía desde ser el cromosoma con el menor valor promedio de  $r^2$  en ventanas  $> 10$ kb a ser el cromosoma con el menor valor promedio de  $r^2$  en ventanas  $< 10$ kb (Figura 3.14). La ordenación de  $r^2$  en los autosomas, sin embargo, es la misma para cualquier tamaño de ventana:  $3R > 2L > 2R = 3L$  (Figura 3.14). Estas observaciones concuerdan con lo observado previamente: a distancias menores de 1kb, el promedio de  $r^2$  del cromosoma X es superior al de los autosomas (Wall *et al.* 2002; Mackay *et al.* 2012; Langley *et al.* 2012), pero a distancias mayores (5-10kb), la  $r^2$  del cromosoma X es comparable a los autosomas y 3R muestra los valores más altos de  $r^2$  (Langley *et al.* 2012).



**Figura 3.14.** Desequilibrio de ligamiento,  $r^2$ , en los bloques de selección ligada (LSB)

Desequilibrio de ligamiento,  $r^2$ , para cada brazo cromosómico (a-d), para los LSB (azul oscuro) y NLSB (azul claro) estimados con  $\rho$  (b-h) y con  $c$  (c-l) en distintos tamaños de ventanas: 100kb (a, e, i), 50kb (b, f, j), 10kb (c, g, k) y 1kb (d, h, l).

**Tabla 3.17.** Diferencias en  $r^2$  entre LSB y NLSB

100kb	LSB		NLSB	
	P <sub>2,5</sub>	P <sub>97,5</sub>	P <sub>2,5</sub>	P <sub>97,5</sub>
2L	0,0190	0,0226	0,0144	0,0156
2R	0,0168	0,0201	0,0119	0,0128
3L	0,0157	0,0176	0,0114	0,0124
3R	0,0225	0,0256	0,0166	0,0191
X	0,0189	0,0246	0,0121	0,0130

50kb	LSB		NLSB	
	P <sub>2,5</sub>	P <sub>97,5</sub>	P <sub>2,5</sub>	P <sub>97,5</sub>
2L	0,0226	0,0270	0,0177	0,0189
2R	0,0212	0,0244	0,0148	0,0159
3L	0,0184	0,0213	0,0144	0,0153
3R	0,0263	0,0291	0,0200	0,0226
X	0,0259	0,0338	0,0158	0,0178

10kb	LSB		NLSB	
	P <sub>2,5</sub>	P <sub>97,5</sub>	P <sub>2,5</sub>	P <sub>97,5</sub>
2L	0,0336	0,0379	0,0285	0,0311
2R	0,0309	0,0347	0,0251	0,0269
3L	0,0285	0,0316	0,0237	0,0255
3R	0,0348	0,0375	0,0327	0,0350
X	0,0411	0,0466	0,0325	0,0359

1kb	LSB		NLSB	
	P <sub>2,5</sub>	P <sub>97,5</sub>	P <sub>2,5</sub>	P <sub>97,5</sub>
2L	0,0630	0,0692	0,0571	0,0627
2R	0,0568	0,0642	0,0520	0,0571
3L	0,0537	0,0592	0,0460	0,0511
3R	0,0650	0,0712	0,0597	0,0660
X	0,0788	0,0949	0,0800	0,0838

Resultado del remuestreo *bootstrap* para los valores de  $r^2$  en los LSB y NLSB. Si los resultados solapan se considera que no hay diferencias significativas. P<sub>2,5</sub> y P<sub>97,5</sub> son el percentil 2,5 y el percentil 97,5 respectivamente.

En ventanas mayores de 10kb, los LSB tienen un valor promedio de  $r^2$  significativamente superior al de los NLSB, según los resultados del remuestreo *bootstrap* (Tabla 3.17). Sin embargo, para tamaños de ventana de 1kb, a pesar de que la tendencia es la misma,  $r^2$  no es significativamente superior en los LSB. Este patrón es similar tanto para  $c$  como para  $\rho$ . Es decir, parece que en ventanas menores de 10kb la resolución a la que se estimó el valor umbral

de recombinación (100kb) no es el adecuado para distinguir dos regiones diferenciadas del genoma en cuanto a la selección ligada.

### 3.6. Correlación entre polimorfismo y divergencia en los LSB

Si como se ha propuesto, los sitios neutros en los NLSB no están afectados por la selección ligada, dichos sitios segregarían según predice la teoría neutralista. El polimorfismo neutro: (1) no debería correlacionarse con la tasa de recombinación (Hudson 1983) y (2) debería estar positivamente correlacionado con la divergencia neutra (Kimura 1983). Por el contrario, en regiones donde actúa la selección ligada se esperaría: (1) una correlación positiva entre el polimorfismo y la recombinación (Begun y Aquadro 1991) y (2) la ausencia de correlación entre polimorfismo y la divergencia. Según la teoría neutralista, tanto  $\theta_s$  ( $4N_e\mu$ ) como  $k_s$  ( $\mu$ ) incrementan con la tasa de mutación, pero bajo la selección ligada la reducción que ésta causa en términos de  $N_e$  impediría observar una correlación entre el polimorfismo neutro y la divergencia.

Con el mapa basado en  $\rho$ , en los LSB hay una correlación fuerte entre la recombinación y el polimorfismo en los sitios 4 veces degenerados (corregido por la divergencia) en ventanas de 100kb. Por otro lado, en estos bloques no se observa una correlación entre el polimorfismo y la divergencia en los sitios 4 veces degenerados. Sin embargo, en los NLSB el patrón se invierte: no se detecta correlación entre la recombinación y el polimorfismo, pero sí que hay una correlación entre el polimorfismo y la divergencia (Tabla 3.18 y Tabla 3.19). Es decir, que en el mapa basado en  $\rho$  se cumplen las predicciones hechas bajo nuestra hipótesis que en los NLSB, la selección ligada, si ocurre, tiene un efecto despreciable.

**Tabla 3.18.** Correlación entre el polimorfismo y la recombinación en los LSB y NLSB

Cromosoma	100kb $\rho$				100kb $c$			
	LSB		NLSB		LSB		NLSB	
	R	valor p	R	valor p	R	valor p	R	valor p
2L	0,6998	$<2,2 \cdot 10^{-16}$	0,2021	0,1065	0,5662	$7,28 \cdot 10^{-14}$	0,1094	0,3433
2R	0,7833	$<2,2 \cdot 10^{-16}$	0,0134	0,9056	0,6687	$<2,2 \cdot 10^{-16}$	0,1377	0,2939
3L	0,7529	$<2,2 \cdot 10^{-16}$	0,2677	0,1681	0,6914	$<2,2 \cdot 10^{-16}$	0,2570	0,0343
3R	0,6865	$<2,2 \cdot 10^{-16}$	0,2089	0,0683	0,5734	$1,5 \cdot 10^{-12}$	0,1973	0,0151
X	0,6337	$9,1 \cdot 10^{-7}$	0,1252	0,1111	0,5839	$1,1 \cdot 10^{-7}$	0,0925	0,2633

Correlación de Spearman entre el polimorfismo corregido por la divergencia en los sitios 4 veces degenerados y las dos medidas de recombinación. Mapas de LSB basados en  $\rho$  y  $c$ .



**Tabla 3.19.** *Correlación del polimorfismo y la divergencia en los LSB y NLSB*

Cromosoma	100kb $\rho$				100kb $c$			
	LSB		NLSB		LSB		NLSB	
	R	valor $p$	R	valor $p$	R	valor $p$	R	valor $p$
2L	-0,0013	0,9874	0,2634	0,0340	0,0920	0,2675	0,0821	0,4779
2R	-0,2961	0,0007	0,4135	0,0001	-0,1465	0,0726	0,4555	0,0003
3L	-0,0914	0,1830	0,4904	0,0088	-0,0400	0,6011	0,2253	0,0629
3R	0,0183	0,7958	0,2464	0,0310	-0,0951	0,2853	0,1834	0,0243
X	-0,0501	0,7155	0,2095	0,0074	0,0535	0,6594	0,1335	0,1057

Correlación de Spearman entre el polimorfismo y la divergencia en los sitios 4 veces degenerados. Mapas de LSB basados en  $\rho$  y  $c$ .

En el mapa basado en  $c$  este patrón general se repite pero con alguna excepción (Tabla 3.18 y Tabla 3.19). En los NLSB del cromosoma 3 se sigue observando una correlación entre la recombinación y el polimorfismo aunque el valor de significación  $p$  sea mucho mayor que en los LSB (valor  $p \sim 0,01$ , en los NLSB, en vez de valor  $p < 10^{-11}$  de los LSB). Además, en algunos casos no se observa una correlación entre el polimorfismo y la divergencia en los NLSB. Se observa, sin embargo, una tendencia general hacia un incremento en la fuerza de la correlación (mayor R de *Spearman*) y una menor significación (menor valor  $p$ ) en las regiones de BSNL comparadas con los LSB. Parece, por lo tanto, que  $\rho$  predice mejor que  $c$  los diferentes bloques a lo largo de los cromosomas. En general, el mapa de selección ligada, aunque consistente con las dos medidas de recombinación, parece ser más fiable el que se ha elaborado a partir del parámetro de recombinación histórica. Esto probablemente se explique por el hecho de que el polimorfismo y la divergencia dependen del tamaño de la población y/o el tiempo. A diferencia de  $c$ , el cual se basa en la tasa de recombinación actual y en unos pocos individuos,  $\rho$ ,  $\theta$  y  $k$  son estimaciones que promedian la fluctuación de la recombinación, la selección y el censo efectivo desde la divergencia entre *D. melanogaster* y *D. yakuba*.

Por último, cabe destacar una observación inesperada: la correlación negativa entre el polimorfismo y la divergencia en los LSB del cromosoma 2R. Se ha descrito previamente una correlación negativa entre  $\theta$  ó  $\pi$  y  $k$  para los sitios no sinónimos y/o sitios distintos a los 4-veces degenerados. Este resultado se ha interpretado bien como una mayor tasa de error en el ensamblaje de regiones no ortólogas en las regiones de baja recombinación (Mackay *et al.* 2012) o como el efecto de la selección positiva en el que a medida que la selección sea más intensa incrementará  $k$  debido al aumento de alelos positivos que se fijan pero se reducirá  $\theta$  debido al incremento de los arrastres por selección positiva (Begun *et al.* 2007; Hahn 2008). Sin embargo, nuestros datos son sitios 4 veces degenerados en regiones ortólogas. Es decir, que o bien las regiones de ortología no están bien caracterizadas o bien las posiciones 4-veces degeneradas están afectadas por la selección. En cualquier caso, se requieren análisis

---

adicionales para cerciorarnos que está señal no es espuria y que tiene un valor biológico, en cuyo caso se tendría que explicar porqué este efecto sólo se observa en el cromosoma 2R.

### 3.7. Eficiencia de selección en los LSB y NLSB

#### 3.7.1. Descripción general de las estimas de selección

Se ha estimado la razón  $k_a/k_s$  en genes codificadores de proteínas usando a *Drosophila yakuba* como taxón externo (sección 2.12). Se utilizó el software DFE-alpha (Keightley y Eyre-Walker 2010) para calcular la proporción de sustituciones adaptativas ( $\alpha$ ) y la tasa relativa de sustituciones adaptativas respecto a las sustituciones neutras ( $\omega_a$ , Schneider *et al.* 2011). De este modo podemos definir un nuevo parámetro  $\omega_d = k_a/k_s - \omega_a$  (comunicación personal de D. Castellano), que estima la tasa relativa de sustituciones ligeramente deletéreas respecto de la tasa de sustitución neutra. Trabajos previos han mostrado la existencia de correlación entre la recombinación y  $k_a/k_s$ , pero no pudieron elucidar si el incremento en  $k_a/k_s$  era debido a un incremento en la sustituciones adaptativas o una reducción de los alelos ligeramente deletéreos (Larracunte *et al.* 2008; Webster y Hurst 2012; Presgrave 2005). Por ello, creemos que obtener las estimas de  $\omega_a$  y  $\omega_d$  ayudará a obtener conclusiones más sólidas.

La proporción de las sustituciones adaptativas ( $\alpha$ ) es del 40% para las regiones codificadoras de los genes ortólogos entre *D. melanogaster* y *D. yakuba*. Este valor es mayor que el estimado (24%) por Mackay *et al.* 2012 (sección 1.2.3 y Tabla 1.4) para este mismo conjunto de datos pero utilizando un método de estimación distinto. Hemos calculado de los datos de estudios previos en *D. melanogaster*, *D. simulans* y *D. yakuba* (en su mayoría con genomas parciales de pocos individuos) el promedio de  $\alpha$  en las regiones codificadoras. El valor es ~48,8%, superior a nuestras estimas (Tabla 1.4). Además, gracias a las estimas de  $\omega_a$  y  $\omega_d$  sabemos que entre los alelos que se fijan ( $k_a/k_s = 0,19$ ), el 49,84% corresponden a alelos adaptativos y el 50,16% a alelos neutros o ligeramente deletéreos. Es decir, que casi la mitad de los alelos que se fijan son adaptativos (Tabla 3.20). Cuando analizamos los resultados por brazos cromosómicos observamos que las diferencias principales están en el cromosoma 3, con una  $\alpha$  y  $k_a/k_s$  ligeramente menor. Entre los cromosomas 2 y X no hay diferencias significativas (Tabla suplementaria 4). Para estos cromosomas, más de la mitad de los alelos que se fijan son adaptativos (~55%), mientras que la mayoría de los que se fijan en el cromosoma 3 son alelos neutros o ligeramente deletéreos (~57%) (Tabla 3.20).

**Tabla 3.20.** *Estimas de selección en el genoma de D. melanogaster*

	100kb				10kb			
	$\alpha$	$k_a/k_s$	$\omega_a$	$\omega_d$	$\alpha$	$k_a/k_s$	$\omega_a$	$\omega_d$
<i>Total</i>	0,4024	0,1904	0,0949	0,0955	0,1006	0,1911	0,0868	0,1041
<i>2L</i>	0,4961	0,1966	0,1093	0,0874	0,3021	0,2072	0,1072	0,0998
<i>2R</i>	0,4092	0,1937	0,0996	0,0942	0,1696	0,1885	0,0834	0,1049
<i>3L</i>	0,3750	0,1707	0,0797	0,0911	-0,0433	0,1835	0,0862	0,0972
<i>3R</i>	0,2606	0,1849	0,0741	0,1108	-0,1102	0,1742	0,0574	0,1166
<i>X</i>	0,5116	0,2090	0,1190	0,0899	0,3006	0,2117	0,1140	0,0972
<i>Autosomas</i>	0,3773	0,1861	0,0894	0,0967	0,0608	0,1870	0,0814	0,1054

Promedios generales para cada brazo cromosómico de  $\alpha$ ,  $\omega_a$ ,  $\omega_d$  y  $k_a/k_s$  estimados en ventanas de 100kb y 10kb.

### 3.7.2. Diferencias entre LSB y NLSB

Si suponemos que la eficiencia de selección incrementa con la recombinación (revisado en Charlesworth 2012), deberíamos observar una mayor eficiencia de selección en los NLSB respecto a los LSB. En los LSB, donde la selección ligada es una fuerza principal, la interferencia de Hill-Robertson (iHR, Hill y Robertson 1968) entre alelos debería extenderse más allá del tamaño de ventana, reduciendo la eficiencia de la selección positiva y la negativa. Por el contrario, en los NLSB no esperamos que la iHR se dé entre ventanas consecutivas, lo que debería incrementar la eficiencia de selección en estas regiones, aunque no podamos excluir la posibilidad de que la iHR se produzca dentro de cada ventana.

En los NLSB se observa un incremento en  $\omega_a$  (~86% con  $\rho$ , ~58% con  $c$ ) y una reducción en  $\omega_d$  (~38% con  $\rho$ , ~32% con  $c$ ) respecto a los LSB (Tabla 3.21). Esto sugiere un incremento en la eficiencia de selección en los NLSB, tanto un incremento en la fijación de alelos adaptativos como una reducción en las sustituciones ligeramente deletéreas. Este incremento en la eficiencia de la selección no se observa cuando se analiza exclusivamente  $k_a/k_s$ , ya que el sentido opuesto del cambio de  $\omega_a$  y  $\omega_d$  se compensan mutuamente haciendo que la razón  $k_a/k_s$  se mantenga inalterada en los LSB y NLSB (excepto para el brazo cromosómico 2R con  $\rho$ ) (Tabla 3.21, Tabla 3.22). Cuando se utiliza  $c$  como medida de recombinación para estimar el valor umbral, las diferencias significativas no se observan en todos los brazos cromosómicos como es el caso del brazo cromosómico 2L (Tabla 3.21, Tabla 3.22). Nuevamente parece que con  $\rho$  se obtienen unas diferencias más nítidas entre LSB y NLSB.

**Tabla 3.21.** *Estimas de selección en los LSB de D. melanogaster*

a)	LSB				NLSB			
	$A$	$k_d/k_s$	$\omega_a$	$\omega_d$	$\alpha$	$k_d/k_s$	$\omega_a$	$\omega_d$
2L	0,4395	0,1948	0,0977	0,0971	0,6266	0,1976	0,1341	0,0635
2R	0,2494	0,1811	0,0721	0,1090	0,6583	0,2165	0,1444	0,0721
3L	0,3607	0,1691	0,0783	0,0908	0,5925	0,1786	0,1200	0,0585
3R	0,1473	0,1836	0,0590	0,1245	0,5395	0,1896	0,1126	0,0770
X	0,1380	0,1992	0,0558	0,1434	0,6182	0,2147	0,1393	0,0753

b)	LSB				NLSB			
	$\alpha$	$k_d/k_s$	$\omega_a$	$\omega_d$	$\alpha$	$k_d/k_s$	$\omega_a$	$\omega_d$
2L	0,4390	0,2030	0,1072	0,0958	0,5932	0,1851	0,1124	0,0727
2R	0,3290	0,2008	0,0957	0,1050	0,5830	0,1785	0,1079	0,0706
3L	0,3233	0,1691	0,0686	0,1005	0,4932	0,1745	0,1050	0,0695
3R	0,04237	0,1776	0,0406	0,1370	0,4464	0,1911	0,1025	0,0886
X	0,2898	0,2028	0,0822	0,1206	0,6171	0,2119	0,1365	0,0753

Promedios para los LSB y NLSB de cada brazo cromosómico de  $\alpha$ ,  $\omega_a$ ,  $\omega_d$  y  $k_d/k_s$  en ventanas de 100kb con (a) el mapa basado en  $\rho$  y (b) el mapa basado en  $c$ .

**Tabla 3.22.** *Diferencias de las estimas de selección en los LSB y NLSB*

a) $\rho$	2L		2R		3L		3R		X	
	W	valor p	W	valor p	W	valor p	W	valor p	W	valor p
$\alpha$	3.905	0,0223	2.760	$1,16 \cdot 10^{-7}$	2.065	0,0714	4.538	$7,22 \cdot 10^{-7}$	2.275	$7,41 \cdot 10^{-6}$
$\omega_a$	4.108	0,0717	2.809	$2,21 \cdot 10^{-7}$	2.098	0,0892	4.496	$4,99 \cdot 10^{-7}$	2.876	0,0041
$\omega_d$	5.889	0,0146	6.543	$9,32 \cdot 10^{-5}$	3.317	0,0332	9.824	$3,14 \cdot 10^{-5}$	5.534	$1,98 \cdot 10^{-5}$
$k_d/k_s$	4.964	0,8125	3.837	0,0074	2.434	0,5202	6.761	0,2632	3.742	0,5876

b) $c$	2L		2R		3L		3R		X	
	W	valor p	W	valor p	W	valor p	W	valor p	W	valor p
$\alpha$	5.044	0,3291	3.179	0,0005	4.395	0,0095	5.842	$1 \cdot 10^{-7}$	3.966	0,0143
$\omega_a$	5.266	0,6323	3.750	0,0364	4.700	0,0527	5.856	$1,13 \cdot 10^{-7}$	4.336	0,1157
$\omega_d$	6.116	0,1544	6.024	0,0003	6.881	0,0058	12.381	$2,94 \cdot 10^{-6}$	6.134	0,0076
$k_d/k_s$	5.804	0,4685	4.865	0,4782	5.893	0,5287	8.646	0,300	4.963	0,9266

Test de Wilcoxon entre los LSB y NLSB para cada uno de los parámetros de selección ( $\alpha$ ,  $\omega_a$ ,  $\omega_d$  y  $k_d/k_s$ ) en ventanas de 100kb con (a) el mapa basado en  $\rho$  y (b) el mapa basado en  $c$ .

### 3.7.3. Comparación entre LSB y NLSB situados fuera de las regiones centroméricas

Las diferencias observadas entre LSB y NLSB podrían deberse exclusivamente al efecto de las regiones centroméricas y teloméricas en la que se ha demostrado ampliamente que la eficiencia de selección es menor (Haddrill *et al.* 2007; McVean y Charlesworth 2000; Bachtrog D 2003). Es decir, que podría no haber diferencias entre LSB y NLSB más allá de los bloques

de los extremos de los brazos cromosómicos. Para probar si las diferencias encontradas entre LSB y NLSB se deben a un mero efecto centromérico, hemos repetido el análisis excluyendo las regiones centroméricas (eliminando el LSB más externo de cada brazo cromosómico, Tabla suplementaria 3).

Aunque las diferencias entre los LSB y los NLSB se reducen, persiste todavía una reducción en  $\omega_d$  (~15%) y un incremento de  $\omega_a$  (~33%) en las NLSB (Tabla 3.23, Tabla 3.24). Pero estas diferencias son sólo significativas en el cromosoma X y para  $\omega_a$  también el brazo cromosómico 2R y 3R (Tabla 3.24). Es decir, que la tendencia parece mantenerse, pero las diferencias desaparecen en los brazos 2L y 3L principalmente. Cuando se analizan los resultados basados en  $c$ , las diferencias se observan únicamente en el brazo 3R. Pero además, la tendencia sólo se mantiene para  $\omega_d$  (~18% menor en los NLSB) y para  $\omega_a$  del cromosoma 3 exclusivamente (~44%).

**Tabla 3.23.** *Estimas de selección en los LSB tras excluir las regiones centroméricas*

	LSB con $\rho$				LSB con $c$			
	$A$	$k_a/k_s$	$\omega_a$	$\omega_d$	$\alpha$	$k_a/k_s$	$\omega_a$	$\omega_d$
2L	0,6437	0,1941	0,1278	0,0663	0,4607	0,2061	0,1130	0,0931
2R	0,5883	0,1901	0,1175	0,0726	0,5606	0,2162	0,1322	0,0840
3L	0,4956	0,1696	0,0986	0,0710	0,4660	0,1740	0,0938	0,0802
3R	0,4696	0,1884	0,1061	0,0823	0,1334	0,1704	0,0581	0,1123
X	0,2372	0,2014	0,0663	0,1351	0,5871	0,2363	0,1445	0,0918

Promedios de cada uno de los parámetros de selección ( $\alpha$ ,  $\omega_a$  y  $\omega_d$ ) en los LSB tras quitar las regiones centroméricas y teloméricas (definidas en la Tabla suplementaria 3). Ventanas de 100kb con el en mapa basado en  $\rho$  y el mapa basado en  $c$ . Nótese que los NLSB no varían respecto a la Tabla 3.21.

Resumiendo, nuestros resultados parecen indicar que los mapas de bloques de selección ligada están delimitando las regiones de alta y baja eficiencia de selección a lo largo de los cromosomas. Sin embargo, no se puede excluir que haya baja eficiencia de la selección fuera de las regiones de nula o baja recombinación. Es decir, que en las regiones centrales de los brazos cromosómicos, no todo NLSB tiene una mayor eficiencia de selección. Este hecho puede que se deba a una falta de potencia estadística, ya que al analizar solamente las regiones centrales se ha disminuido el tamaño y conjunto de datos de forma considerable.

**Tabla 3.24.** Diferencias de las estimas de selección en los LSB y NLSB tras excluir las regiones centroméricas

a) $\rho$	2L		2R		3L		3R		X	
	W	valor p	W	valor p	W	valor p	W	valor p	W	valor p
$\alpha$	2.923	0,7829	1.733	0,0947	1.661	0,4172	4.538	0,1370	1.853	$1,49 \cdot 10^{-5}$
$\omega_a$	2.931	0,7603	1.619	0,0278	1.678	0,4626	4.496	0,0456	2.333	0,0038
$\omega_d$	2.913	0,8114	2.174	0,7104	2.133	0,2098	9.824	0,4879	4.718	$2,23 \cdot 10^{-5}$

b) c	2L		2R		3L		3R		X	
	W	valor p	W	valor p	W	valor p	W	valor p	W	valor p
$\alpha$	4.878	0,6311	2.594	0,3271	3.823	0,2597	4.063	0,0001	2.902	0,8021
$\omega_a$	5.075	0,9914	2.912	0,8492	4.026	0,7741	3.951	$5,83 \cdot 10^{-5}$	3.229	0,1744
$\omega_d$	5.548	0,2653	3.236	0,1658	4.664	0,1391	7.160	0,0053	3.116	0,3293

Test de Wilcoxon entre los LSB y NLSB para cada uno de los parámetros de selección ( $\alpha$ ,  $\omega_a$  y  $\omega_d$ ) en ventanas de 100kb con el mapa basado en  $\rho$  (a) y el mapa basado en  $c$  (b) tras quitar las regiones centroméricas y teloméricas (definidas en la Tabla suplementaria 3).

### 3.8. Mapa de LSB para diversos tamaños de ventana

De la definición de LSB se infiere que el tamaño de ventana es una variable crucial para la detección de los LSB (sección 3.3.3). Para probar el efecto que el tamaño de ventana puede tener en la determinación de los LSB hemos calculado el parámetro de recombinación ( $\rho$ ) y el polimorfismo ( $\theta$ ) en diferentes tamaños de ventana además de 100kb: 1Mb, 10kb y 1kb.

**Tabla 3.25.** Relación entre el polimorfismo y la recombinación para distintos tamaños de ventana

Tamaño de ventana	$\rho$ por $p_b$	$\theta$	Correlación de Pearson
1Mb	0,0197	0,0103	$r = 0,5619$ $valor p = 2 \cdot 10^{-11}$
100kb	0,0110	0,0105	$r = 0,4620$ $valor p = 2 \cdot 10^{-16}$
10kb	0,0211	0,0091	$r = 0,3643$ $valor p = 2 \cdot 10^{-16}$
1kb	0,0356	0,0087	$r = -0,0520$ $valor p = 2 \cdot 10^{-16}$

Valores medios de polimorfismo y correlación de Pearson entre el polimorfismo y el parámetro de recombinación poblacional ( $\rho$ ) en ventanas de 1Mb, 100kb, 10kb y 1kb.

El polimorfismo por sitio promedio de cada brazo cromosómico es similar, independientemente del tamaño de ventana que se utilice. Sin embargo, con  $\rho$  no ocurre lo mismo. Para reducir el coste computacional, se utilizó una metodología ligeramente diferente para estimar  $\rho$  en estos nuevos tamaños de ventana (se redujo el número de iteraciones, véase Materiales y Métodos sección 2.4). Esta aproximación incrementó en casi 2 veces el parámetro

---

de recombinación por sitio en ventanas de 1Mb y 10kb comparado a las estimas de 100kb (Tabla 3.25). En ventanas de 1kb hay un incremento de  $\sim 1,7$  en  $\rho$  que no se puede explicar por el hecho de utilizar una metodología diferente a las estimaciones de las ventanas de 100kb. El cambio principal se produce en las regiones centroméricas y teloméricas donde en vez de observarse una reducción de la recombinación (como se observa para el polimorfismo y también para el parámetro de recombinación a cualquier otro tamaño de ventana) se observa un incremento en  $\rho$ . Por eso, a este tamaño de ventana, la correlación positiva entre el polimorfismo y la recombinación desaparece (Tabla 3.25). En consecuencia, la correlación característica observada entre el polimorfismo y la recombinación sólo se observa en ventanas  $\geq$  a 10kb. Por debajo de este tamaño de ventana, la correlación entre el parámetro de recombinación y el polimorfismo es negativa (Tabla 3.25). Tener mayor número de ventanas con pocos sitios segregantes (como puede ser el caso de las ventanas a 1kb) puede hacer fluctuar a las estimas de polimorfismo y de  $\rho$  de tal forma que se reduzca o se pierda la correlación entre ambas. Sin embargo, es difícil suponer que este incremento en la variación estocástica de los datos genere una correlación negativa entre el polimorfismo y la recombinación. Interpretamos pues que nuestras estimas del parámetro de recombinación están incrementadas en ventanas de 1kb debido probablemente al papel de la conversión génica. Cómo se argumentará en detalle en la Discusión (sección 4.5),  $\rho$  es un estimador del entrecruzamiento en ventanas mayores, pero en ventanas de pequeño tamaño, la conversión génica adquiere un mayor protagonismo (Berry y Barbadilla 2000; Andolfatto y Nordborg 1998; Langley *et al.* 2000; Frisse *et al.* 2001) y  $\rho$  acaba siendo un parámetro afectado tanto por el entrecruzamiento como por la conversión génica. En ventanas de 1kb,  $\rho$  estaría completamente gobernado por la conversión génica, y la conversión limita la observación de una correlación positiva entre recombinación y polimorfismo. Se ha descrito recientemente una correlación negativa a lo largo del genoma entre la tasa de entrecruzamiento y la tasa de conversión génica en *D. melanogaster* (Comeron *et al.* 2012, Cuadro 3). Según ello, si el polimorfismo está positivamente correlacionado con el entrecruzamiento, cabe esperar una correlación negativa entre el polimorfismo y la conversión génica, como la que se observa en nuestros resultados. Además de las consecuencias evolutivas que esto pueda suponer, a la práctica nos imposibilita estimar de forma precisa el entrecruzamiento en ventanas menores a 10kb y esto hace que no podamos estimar el valor umbral de recombinación en estas ventanas. Por tanto, no podremos detectar si el valor umbral de recombinación se mantiene inalterado cuando disminuimos el tamaño de la ventana.

### 3.8.1. El mapa de LSB y NLSB

Hemos construido mapas de LSB adicionales en ventanas de 1Mb y 10kb (Figura 3.15). En autosomas, el valor umbral de recombinación se incrementa  $\sim 1,4$  veces desde el tamaño de ventana de 1Mb a 10kb, lo que conlleva un incremento de un 18% de la proporción de sitios que

---

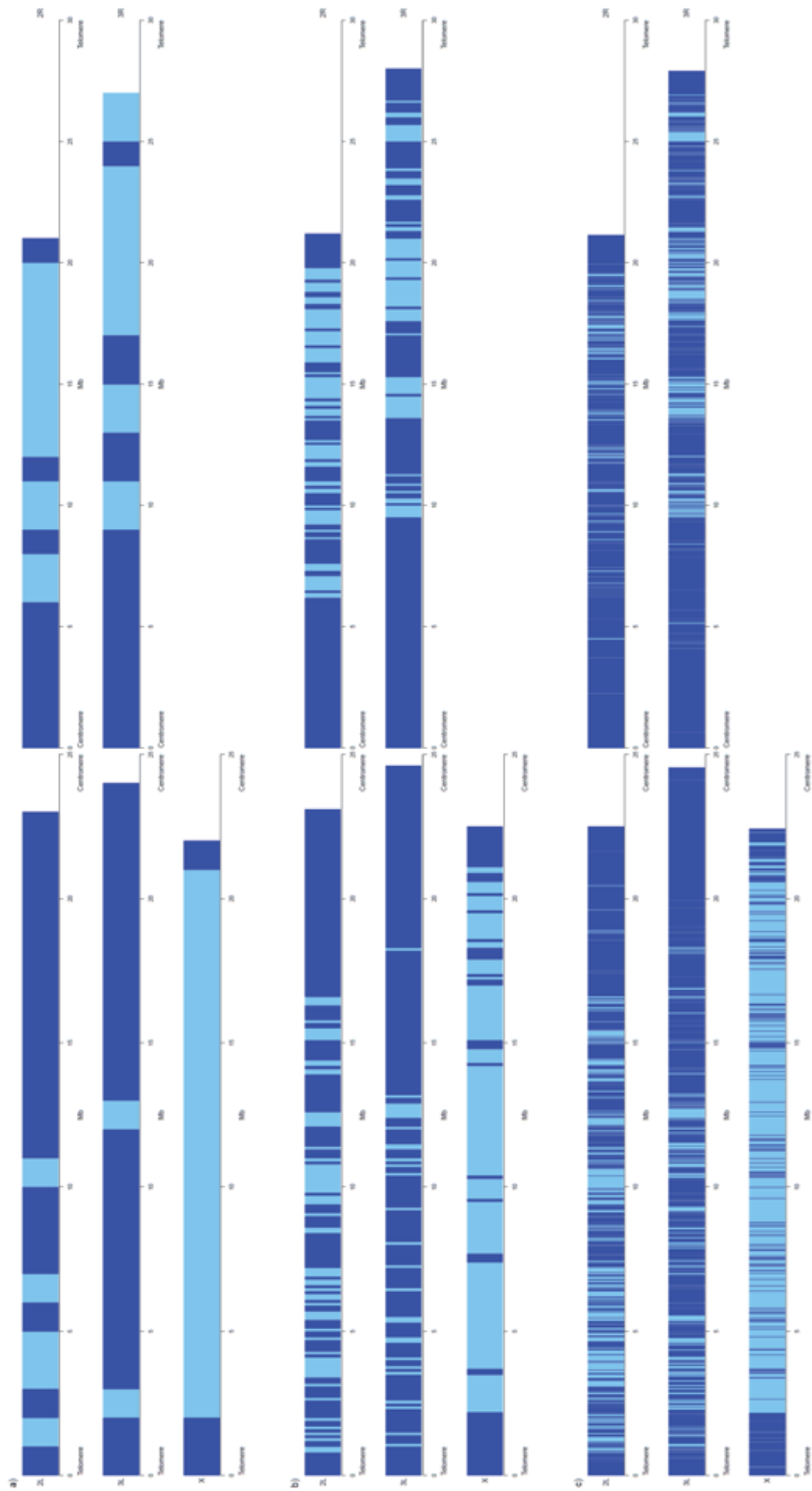
forman los LSB (11% del mapa de 1Mb al mapa de 100kb y otro incremento del 7% del mapa de 100kb al mapa de 10kb). Sin embargo, en el cromosoma X, a pesar de que el valor umbral de recombinación disminuye 1,7 veces, la proporción de sitios que forman los LSB incrementan en solo un 9% (5% del mapa de 1Mb al de 100kb y un 4% más del mapa de 100kb al de 10kb) (Tabla 3.10, Tabla 3.13, Tabla suplementaria 2). Esto pone de manifiesto que el valor umbral de recombinación en sí mismo puede ser muy variable en función de la ventana que se utilice y parece ser más informativo comparar la proporción de sitios que hay por encima y por debajo de ese valor umbral (la % de sitios o nucleótidos que ocupan los LSB y NLSB) para comparar resultados en distintos tamaños de ventanas. A pesar de toda esta variabilidad, el patrón general es consistente para cualquier tamaño de ventana. En autosomas, observamos una mayor proporción de sitios que forman LSB comparado con los NLSB (LSB 1Mb: 62%; 100kb: 73%; 10kb: 80%) pero en el cromosoma X se observa el patrón opuesto (LSB 1Mb: 17%; 100kb: 23%; 10kb: 27%). El tamaño promedio de los LSB y los NLSB disminuyen con el tamaño de ventana (~7 veces menor de 1Mb al mapa de 100kb y ~10 veces menor al mapa 10kb; Tabla 3.12). Pero al igual que en el mapa de 100kb, para autosomas los LSB son en promedio más largos que los NLSB (~3 veces para el mapa de 1Mb a 100kb y ~4 veces para el mapa de 10kb). Para el cromosoma X, sin embargo, en cualquier tamaño de ventana los NLSB son mayores que los LSB, pero las diferencias son menores a menor tamaño de ventana (NLSB son 1Mb: x12,7; 100kb: x2,9; 10kb: x2,1).

### **3.8.2. Evaluación del mapa de LSB en ventanas de 10kb**

#### **3.8.2.1. Correlación entre el polimorfismo y la divergencia**

En ventanas de 10kb, al igual que a 100kb, la correlación entre el polimorfismo y la recombinación desaparece en los NLSB en la mayoría de los casos (Tabla 3.26). Sin embargo, para 10kb la correlación entre el polimorfismo y la divergencia es significativa tanto en los LSB como en los NLSB (Tabla 3.26). Ciertamente es que en los LSB el valor de la correlación y su significación es menor, pero en la mayoría de los casos sigue siendo significativo. Es decir, que las predicciones de la teoría neutralista se cumplen en los NLSB, pero también en los LSB respecto al polimorfismo y la divergencia. Por tanto, a este tamaño de ventana, parece que no se consiguen delimitar los bloques de forma precisa. El método de estimación del umbral parece ser demasiado conservador en el mapa de 10kb e incluye también regiones no afectadas por la selección ligada dentro de los LSB.





**Figura 3.15.** Mapa de LSB en diferentes tamaños de ventana

Mapa de LSB basado en el parámetro poblacional de recombinación  $\rho$ , para cada brazo cromosómico y distintos tamaños de ventana; (a) 1MB, (b) 100kb y (c) 10kb. LSB = azul claro y NLSB = azul oscuro.

**Tabla 3.26.** Correlaciones polimorfismo-recombinación y polimorfismo-divergencia en los LSB en ventanas de 10kb

Cromosoma	$\theta/k - \rho$				$\theta - k$			
	LSB	NLSB	LSB	NLSB	LSB	NLSB	LSB	NLSB
	R	valor p	R	valor p	R	valor p	R	valor p
2L	0,5196	$<2,2 \cdot 10^{-16}$	0,0708	0,1309	0,1171	0,0001	0,3428	$7,1 \cdot 10^{-15}$
2R	0,5678	$<2,2 \cdot 10^{-16}$	0,1481	0,0359	0,1515	$2,2 \cdot 10^{-8}$	0,4090	$5,4 \cdot 10^{-10}$
3L	0,6130	$<2,2 \cdot 10^{-16}$	0,1828	0,0043	0,1272	$1,5 \cdot 10^{-6}$	0,4308	$2,6 \cdot 10^{-13}$
3R	0,5691	$<2,2 \cdot 10^{-16}$	0,1331	0,0093	0,1663	$1,6 \cdot 10^{-11}$	0,3527	$3,7 \cdot 10^{-13}$
X	0,3724	$1 \cdot 10^{-13}$	0,0077	0,8071	0,1828	$6,9 \cdot 10^{-5}$	0,2193	$4,8 \cdot 10^{-13}$

Correlación de Spearman entre el polimorfismo corregido por la divergencia en los sitios 4 veces degenerados y el parámetro poblacional de recombinación ( $\theta/k - \rho$ ). Correlación de Spearman entre el polimorfismo y la divergencia de los sitios 4 veces degenerados ( $\theta - k$ ) en los LSB y los NLSB en ventanas de 10kb.

### 3.8.2.2. Eficiencia de selección

Los valores promedios generales de las estimas de selección por brazo cromosómico y distintos tamaños de ventanas son similares a los promedios obtenidos a 100kb excepto para la variable  $\alpha$  (Tabla 3.20 y Tabla 3.27). Los valores de  $\alpha$  parece que se reducen considerablemente cuando el tamaño de ventana utilizado es menor, de un 40% a tan solo el 10%. Este hecho parece indicar que  $\alpha$  no es un estimador robusto y que es difícil comparar de forma precisa distintos estudios con metodologías diferentes. Sin embargo, las estimas de  $\omega_a$  y  $\omega_d$  parecen más robustas puesto que siguen indicando que prácticamente la mitad de los alelos que se fijan son adaptativos (45%) y la otra mitad neutros o ligeramente deletéreos (55%), independientemente del tamaño de ventana utilizado (Tabla 3.20).

Cuando comparamos los LSB *versus* los NLSB en ventanas de 10kb hay diferencias significativas en todos los brazos cromosómicos para  $\omega_a$  (Tabla 3.27). Es decir, que la selección adaptativa parece más eficiente en los NLSB tal y como predice la teoría. Sin embargo, la eficiencia de la selección negativa no muestra un patrón tan claro. Aunque la tendencia a un menor  $\omega_d$  (mayor eficiencia de la selección negativa) se observa en los NLSB de todos los brazos cromosómicos, estas diferencias sólo son significativas en los brazos 3R y X (Tabla 3.27). Por tanto, podemos concluir que nuestro mapa a 10kb estima bien las regiones del genoma con baja eficiencia de selección (LSB) y alta eficiencia de selección (NLSB) en función de la relevancia de la selección ligada en unas y otras regiones.

Quisimos probar el efecto que las regiones centroméricas y teloméricas tienen en las estimas de selección en los LSB y NLSB. Siguiendo el mismo procedimiento que en ventanas de 100kb, quitamos el LSB más cercano al centrómero y telómero (Tabla suplementaria 3). En el mapa de 10kb los LSB eliminados son tan pequeños que no hay diferencias significativas

entre el mapa de LSB completo y los LSB tras excluir los LSB centromérico y telomérico (valor  $p > 0,58$  en todos los brazos cromosómicos y para todos los estimadores). Por tanto, en el mapa de 10kb no observamos diferencias a lo descrito anteriormente (Tabla 3.28) y parece que el efecto se mantiene más allá de las regiones exclusivamente centroméricas y teloméricas.

**Tabla 3.27.** *Estimas de selección en los LSB en ventanas de 10kb*

a)	LSB				NLSB			
	$\alpha$	$k_a/k_s$	$\omega_a$	$\omega_d$	$\alpha$	$k_a/k_s$	$\omega_a$	$\omega_d$
2L	0,2238	0,2024	0,0976	0,1048	0,5155	0,2185	0,1350	0,0835
2R	0,1377	0,1779	0,0740	0,1039	0,3793	0,2554	0,1464	0,1090
3L	-0,1170	0,1761	0,0767	0,0994	0,4028	0,2257	0,1413	0,0844
3R	-0,1949	0,1670	0,0444	0,1226	0,2568	0,2064	0,1159	0,0905
X	-0,1136	0,2106	0,0698	0,1408	0,4820	0,2146	0,1328	0,0818

b)	2L		2R		3L		3R		X	
	W	valor p	W	valor p	W	valor p	W	valor p	W	valor p
$\alpha$	105.975	0,0852	64.027	0,2950	80.425	0,0401	132.868	$1,84 \cdot 10^{-6}$	71.014	0,0001
$\omega_a$	98.861	0,0010	54.630	0,0001	76.240	0,0024	133.477	$2,98 \cdot 10^{-6}$	70.219	$5,45 \cdot 10^{-5}$
$\omega_d$	118.955	0,2202	65.107	0,3593	92.630	0,3923	175.123	0,0012	96.263	0,0008

(a) Promedios de  $\alpha$ ,  $\omega_a$ ,  $\omega_d$  y  $k_a/k_s$  de cada brazo cromosómico para los LSB y NLSB estimados con el mapa basado en 10kb. (b) Prueba de Wilcoxon entre los LSB y NLSB para cada uno de los parámetros de selección ( $\alpha$ ,  $\omega_a$  y  $\omega_d$ ) en ventanas de 10kb.

**Tabla 3.28.** *Estimas de selección en los LSB en ventanas de 10kb tras excluir las regiones centroméricas*

a)	LSB con $p$		
	$\alpha$	$\omega_a$	$\omega_d$
2L	0,2571	0,1021	0,1034
2R	0,1649	0,0787	0,0988
3L	-0,1194	0,0764	0,0994
3R	-0,2030	0,0449	0,1207
X	-0,0382	0,0816	0,1344

b)	2L		2R		3L		3R		X	
	W	valor p	W	valor p	W	valor p	W	valor p	W	valor p
$\alpha$	97.514	0,0818	60.781	0,3372	80.122	0,0402	123.776	$9,3 \cdot 10^{-7}$	67.208	0,0004
$\omega_a$	91.942	0,0022	51.531	0,0002	75.859	0,0022	124.451	$1,7 \cdot 10^{-6}$	67.058	0,0004
$\omega_d$	109.630	0,2243	59.615	0,2049	92.261	0,3944	171.655	0,0009	89.524	0,0019

(a) Promedios de LSB de cada uno de los parámetros de selección ( $\alpha$ ,  $\omega_a$  y  $\omega_d$ ) tras quitar las regiones centroméricas y teloméricas (definidas en la Tabla suplementaria 3) en ventanas de 10kb. (b) Test de Wilcoxon entre los LSB y NLSB para cada uno de los parámetros de selección ( $\alpha$ ,  $\omega_a$  y  $\omega_d$ ) tras quitar las regiones centroméricas y teloméricas en ventanas de 10kb.



---

## Parte 4

## Discusión



---

## 4. Discusión

### 4.1. Genómica de poblaciones en *Drosophila melanogaster*

La disponibilidad de 158 genomas de alta calidad, con una alta cobertura, de una población natural de la especie *Drosophila melanogaster* nos ha brindado una oportunidad sin precedentes para realizar el estudio más exhaustivo en genética de poblaciones llevado a cabo hasta el momento en cualquier especie. La descripción detallada de la variación nucleotídica en todo el genoma obtenido por primera vez en este análisis supone un valor en sí mismo. Estudios previos en genómica de poblaciones en *Drosophila simulans* (Begun *et al.* 2007) y *Drosophila melanogaster* (Sackton *et al.* 2009) se basaron en secuencias de baja cobertura con un número de 7 y 6 líneas, respectivamente, valores manifiestamente insuficientes para todo estudio de genética de poblaciones cuyas inferencias se basen en las frecuencias de las variantes. Una visión genómica, y por lo tanto global, de la diversidad permite plantearse preguntas de la genética de poblaciones cuyas respuestas en estudios previos eran incompletas, e incluso contradictorias, debido a que se basaban en muestras genéticas fragmentarias y no aleatorias. La descripción de la variación en el nivel genómico permite controlar efectos locales *versus* regionales a lo largo de los brazos cromosómicos y de este modo dilucidar el papel que las distintas fuerzas evolutivas juegan en la modulación de la diversidad nucleotídica en distintas regiones del genoma.

En este trabajo se han descrito y analizado 158 líneas de *D. melanogaster*, junto a la secuencia genómica de la especie cercana *D. yakuba*, mediante una batería de métodos comparativos basados en el análisis del polimorfismo y la divergencia para responder a preguntas de interés fundamental en la genómica de poblaciones: ¿Que patrón o gradientes sigue la variación genética a lo largo del genoma? ¿Cuál es el efecto de la recombinación sobre la variación nucleotídica y la evolución molecular a lo largo del genoma? Algunos de nuestros resultados reiteran patrones de variación nucleotídica previamente observados en *D. melanogaster* en conjuntos de datos menores (Andolfatto 2005; Sackton *et al.* 2009), pero aquí se redefinen, mejoran y en algunos casos se esclarecen debido a que se analizan secuencias a escala genómica, para un gran número de genomas individuales y de alta calidad gracias a una cobertura elevada. Por ejemplo, se han obtenido estimas precisas de diversidad nucleotídica de las distintas clases funcionales del genoma que permiten reafirmar las observaciones previas en distintos trabajos respecto al orden relativo de los valores de diversidad y de la naturaleza de la selección en estas clases de sitios (Halligan 2004; Haddrill *et al.* 2005a; Andolfatto 2005). Por otro lado, este análisis también ha permitido esclarecer cuestiones controvertidas como la hipótesis de la evolución rápida del X que propone que el cromosoma X evoluciona a una tasa mayor que los autosomas (Charlesworth *et al.* 1987). Estudios previos que trataban de probar esta hipótesis en *Drosophila* habían llegado a conclusiones contradictorias (Thornton *et al.*

---

2006; Connallon 2007; Singh *et al.* 2007; Baines *et al.* 2008). Debido a que los niveles de variación nucleotídica pueden estar afectados por la posición en el cromosoma o los niveles de recombinación, las respuestas a esta y otras preguntas no se pueden abordar adecuadamente si estas fuentes potenciales de sesgo no son controladas, es decir, sin un análisis completo del genoma. La heterogeneidad en las tasas de evolución de distintas regiones de los brazos cromosómicos podría explicar los resultados discrepantes hallados previamente (véase sección 4.8).

## 4.2. La recombinación es el principal factor modulador de la variación nucleotídica a lo largo de los cromosomas

La caracterización exhaustiva de decenas de líneas nos ha permitido centrarnos en uno de los factores genómicos cuya relevancia parece ser primordial para describir los patrones de variación observados: la recombinación en el nivel de DNA. A pesar de que otras variables como la tasa de mutación juegan su papel, la recombinación es sin duda la variable principal a la hora de explicar el patrón de polimorfismo observado a lo largo de los brazos cromosómicos. Aunque previamente se había propuesto la recombinación como factor principal modulador de la variación en el genoma (Sackton *et al.* 2009), este trabajo confirma y refuerza la magnitud de su efecto. La recombinación explica entre tres (en autosomas) y 8 veces (cromosoma X) más la variación polimórfica observada que lo que explica la tasa de mutación u otras variables analizadas. Cuando un alelo es seleccionado hasta su fijación (por ser adaptativo) o eliminación (por ser deletéreo) en la población, la variación neutra colindante, ligada, se reduce. Cuanto menor es la recombinación mayor son estas regiones ligadas y mayor la reducción del polimorfismo. Este efecto de arrastre parece ser la explicación a la correlación que se ha encontrado reiteradamente entre la recombinación y el polimorfismo en *Drosophila* (Begun y Aquadro 1992, Begun *et al.* 2007, Kulathinal *et al.* 2008, Sackton *et al.* 2009, Stevison y Noor 2010).

La teoría y observaciones previas apuntan a que estas regiones de baja recombinación sufren una disminución de la eficiencia de selección (Charlesworth y Charlesworth 2000; McVean y Charlesworth 2000; Bachtrog y Charlesworth 2002; Bachtrog 2003; Bachtrog 2005; Bartolomé y Charlesworth 2006; Haddrill *et al.* 2007), lo que permitiría que un mayor número de alelos ligeramente deletéreos se fijen por deriva. La reducción del polimorfismo en torno a los centrómeros (Aguadé *et al.* 1989) y la correlación entre polimorfismo y recombinación (revisado en Smukowski y Noor 2011; Cutter y Payseur 2013) se ha reportado repetidamente en la literatura así como la reducción de la eficiencia de la selección en regiones o cromosomas sin recombinación (Charlesworth y Charlesworth 2000; McVean y Charlesworth 2000; Bachtrog and Charlesworth 2002; Bachtrog 2003; Bachtrog 2005; Bartolomé y Charlesworth 2006; Haddrill *et al.* 2007), pero un estudio de esta correlación a nivel genómico no se había podido



---

analizar hasta este trabajo. La reducción de la diversidad nucleotídica causada por la reducción de la recombinación se observa en al menos un tercio de la longitud de cada brazo cromosómico: las regiones que hemos denominados centroméricas. Estas regiones no se circunscriben exclusivamente al centrómero o a regiones con recombinación nula, como se había supuesto inicialmente, sino que abarcan regiones recombinantes más extensas.

Además, en estas regiones centroméricas también se observa un incremento de la divergencia en las posiciones 4 veces degeneradas, lo que podría indicar un incremento de la tasa de mutación. Durante mucho tiempo se ha considerado que las mutaciones sinónimas no tienen un efecto funcional, es decir, que son *silenciosas* y por tanto selectivamente neutras (Kimura 1968; King y Jukes 1969). Teniendo esto en cuenta, la variación sinónima se ha considerado como variación neutra en numerosos estudios de genética de poblaciones y la tasa de sustitución equivalente a la tasa de mutación (McDonal y Kreitman 1991; Sawyer y Hartl 1992; Bustamante *et al.* 2005; Eyre-Walker y Keightley 1999; Yang 2007; Mackay *et al.* 2012). La observación del sesgo en el uso de codón en muchos organismos fue el primer indicio de un posible papel funcional de los sitios sinónimos (Ikemura 1981; Grantham 1981; Gouy y Gautier 1982; Sharp y Devine 1989; Akashi y Eyre-Walker 1998; Plotkin y Kudla 2010; Hershberg y Petrov 2009). Aunque este efecto está ampliamente aceptado en general, suele considerarse un efecto débil o casi neutro (Yang y Nielsen 2008; Hartl *et al.* 1994; Andolfatto *et al.* 2011; Clemente y Vogl 2012; Vogl y Clemente 2012; Zeng y Charlesworth 2009; Zeng y Charlesworth 2010; Chamary *et al.* 2006; Nielsen *et al.* 2007; Akashi 1995). De hecho, los sitios sinónimos parecen evolucionar de forma más lenta de lo esperado bajo neutralidad en muchos organismos, ajustándose mejor a un modelo de acción de la selección débil (Yang y Nielsen 2008; Chamary *et al.* 2006; Nielsen *et al.* 2007; Akashi 1995; Singh *et al.* 2007; Akashi 1996; Eory *et al.* 2010; Künstner *et al.* 2011; Pollard *et al.* 2010; McVean y Viera 2001; Zhou *et al.* 2010; Haddrill *et al.* 2011). No obstante, recientemente se ha estimado que el 22% de las mutaciones sinónimas en las posiciones 4 veces degeneradas son suficientemente deletéreas como para desaparecer rápidamente del genoma de *D. melanogaster* y dejar de segregarse en la población, lo que sugiere un papel funcional de las posiciones 4 veces degeneradas (Lawrie *et al.* 2013). El estudio se realizó en 130 líneas de DGRP y se controló tanto la demografía como la tasa de mutación. Asimismo, se ha descrito que los intrones cortos (<86pb) son más variables que las posiciones 4 veces degeneradas, lo que parece indicar que las posiciones que se aproximan más a un patrón de evolución neutro en el genoma son las posiciones de los intrones cortos y no la de las 4 veces degeneradas (Haddrill *et al.* 2005a; Singh *et al.* 2009; Parsch *et al.* 2010; Clemente y Vogl 2012). En esta tesis no se analizaron los intrones cortos por separado y eso podría explicar que la categoría de intrones en su conjunto sea menos variable que las posiciones 4 veces degeneradas. En resumen, cada vez parece más evidente que los cambios

---

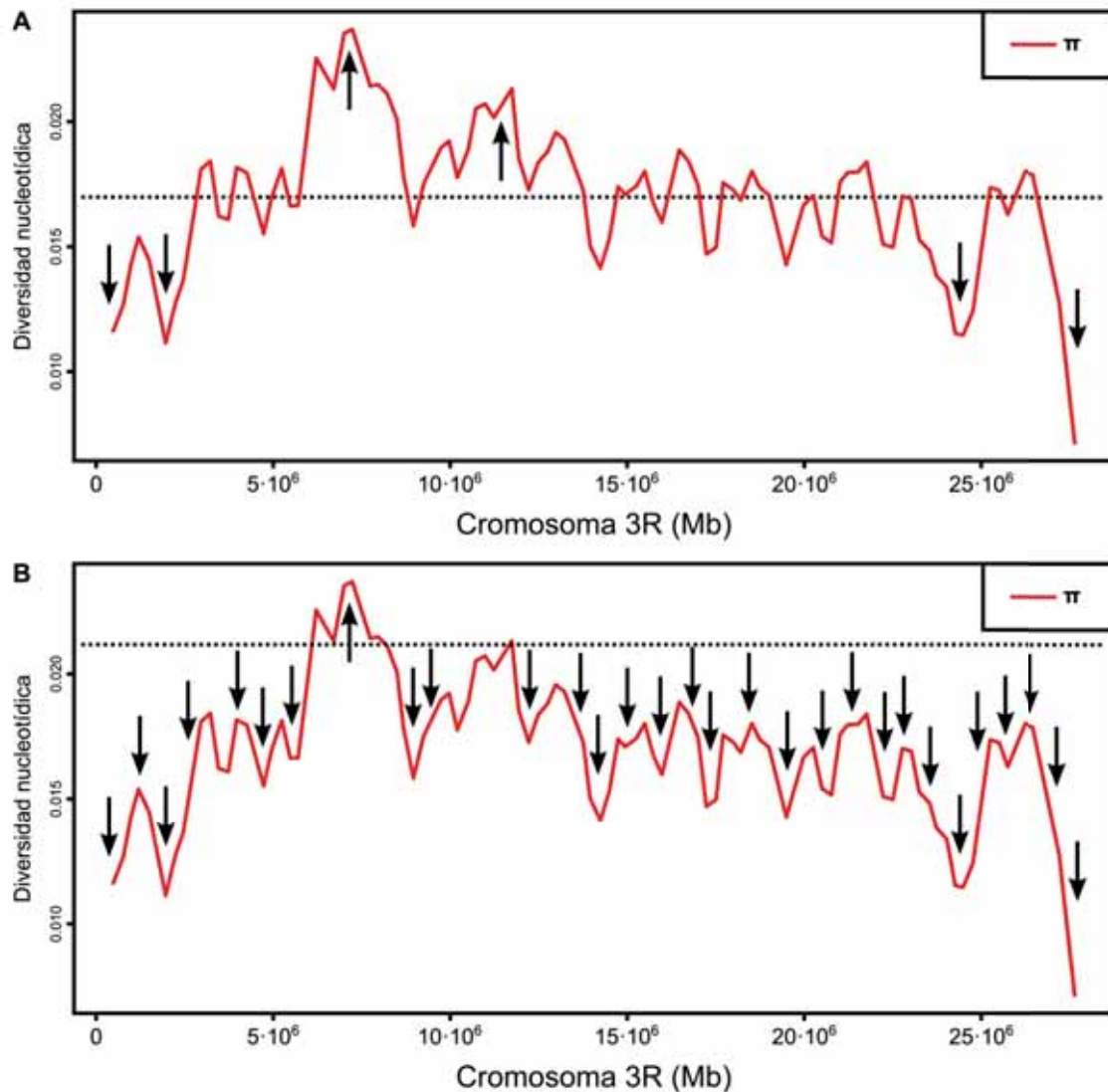
sinónimos que incluyen las posiciones 4 veces degeneradas están sujetos a la acción de la selección purificadora. Si así fuera, la disminución de la eficiencia de selección en las regiones centroméricas haría que la presión selectiva disminuyera respecto a las otras regiones cromosómicas, permitiendo que las posiciones 4 veces degeneradas segregaran como si fueran prácticamente neutras. Esto contrastaría con las regiones selectivamente más eficientes, en las que la selección purificadora podría reducir la fijación de alelos en las posiciones 4 veces degeneradas. Es decir, que la reducción de la eficiencia de la selección observada en la región centromérica podría explicar también el incremento de la fijación de alelos en las posiciones 4 veces degeneradas que están sujetas a selección deletérea. Aunque ni la hipótesis del incremento de la tasa de mutación ni la de la disminución de la eficiencia de selección se pueden rechazar a partir de los análisis realizados, es difícil suponer un mecanismo que incremente la tasa de mutación globalmente en un tercio del cromosoma. Por tanto, parece una explicación más plausible que la eficiencia de la selección disminuya en esta región debido al impacto que supone la disminución de la recombinación, resultando en un incremento de la divergencia en todas las clases de sitios, incluyendo las posiciones 4 veces degeneradas.

### 4.3. Bloques de selección ligada: definición, medida e implicaciones

El mecanismo que subyace a la selección ligada, sus implicaciones generales y efectos locales han sido previamente descritos en varios trabajos (Smith y Haigh 1974; Begun y Aquadro 1992; Charlesworth *et al.* 1993; McVean y Charlesworth 2000; Comeron y Kreitman 2002; Hadany y Comeron 2008; Hahn 2008; Good *et al.* 2014). La relevancia de la selección ligada en la evolución genómica se puso de manifiesto tras la observación de la correlación entre la recombinación y el polimorfismo en un abanico muy amplio de especies (revisado en Smukowski y Noor 2011; Cutter y Payseur 2013). Basándose en esta correlación, se suele suponer implícitamente que la selección ligada afecta prácticamente al genoma completo de *Drosophila* (Hahn 2008; Hadany y Comeron 2008). Sin embargo, no hay ningún trabajo que haya explorado esta idea en el nivel genómico y consecuentemente la selección ligada sigue siendo un concepto cualitativo e impreciso. En esta tesis se ha intentado clarificar y concretar el efecto de la selección ligada a nivel genómico, introduciendo el concepto de Bloque de Selección Ligada (LSB).

Por primera vez se ha conseguido cartografiar aquellas regiones genómicas en las que la selección ligada es una fuerza determinante de los patrones de variación y aquellas regiones en las que la selección ligada es una fuerza menor. Se ha cuantificado el tamaño y la proporción de los bloques de selección ligada a lo largo del genoma. Según nuestros resultados, el polimorfismo y la recombinación no se correlacionaron en un 27% de los autosomas y un 77% del cromosoma X. Es decir, en estas regiones, que comprenden un 40% del genoma, la importancia de la selección ligada como fuerza evolutiva es despreciable. En estas últimas

regiones sin selección ligada, los patrones de variación observados coincidirían con los que predice la teoría neutralista de la evolución molecular (Kimura 1983).



**Figura 4.1.** Interpretación de los niveles de diversidad nucleotídica bajo (a) la teoría neutralista y (b) la selección ligada recurrente (draft genético)

La línea roja representa valores de polimorfismo ( $\pi$ ) a lo largo del cromosoma 3R de *D. simulans* (Begun *et al.* 2007). Las flechas indican el efecto hipotético de la selección ligada incrementando o reduciendo los niveles de polimorfismo. La línea discontinua representa los niveles de polimorfismo bajo neutralidad, dado como el valor medio de  $\pi$  en el cromosoma 3R en el panel A y como el hipotético valor no afectado por la selección ligada en el panel B. Gráfico tomado de Hahn 2008.

La implicación principal de que la selección ligada sea omnipresente en los genomas es la visión distorsionada que se ha tenido de la evolución molecular al considerar cada nucleótido como una unidad de selección (Bustamante *et al.* 2001; Hahn 2008; Neher 2013; Messer y Petrov 2013). Si los modelos en los que se basan las hipótesis nulas de la genética de poblaciones suponen la ausencia de la selección ligada, y ésta por el contrario es ubicua, entonces los métodos de inferencia demográfica y de detección y medida de la selección natural

---

son inadecuados (Figura 4.1). En las regiones donde la selección ligada es una fuerza secundaria o inexistente, como en los NLSB, la interpretación tradicional o clásica de los datos de variación basados en la neutralidad como modelo nulo podría adoptarse perfectamente (Cavalli-Sforza 1966; Lewontin y Krakauer 1973). Bajo una visión neutralista, la media de  $\pi$  representa el polimorfismo esperado y las desviaciones más extremas de esta media indicarían la acción de la selección equilibradora (valores elevados) o selección positiva (valores reducidos) (Figura 4.1a). Sin embargo, en regiones donde la selección ligada es una fuerza evolutiva principal (como en los LSB) los *loci* exhiben niveles de polimorfismo muy por debajo de los esperados bajo la teoría neutralista (Figura 4.1b). En este caso, los genes con niveles elevados de polimorfismo serán los que escapen del efecto de la selección ligada o estén afectados por la selección equilibradora. Estos dos escenarios de la teoría evolutiva llevan a conclusiones completamente diferentes sobre la cantidad de selección que actúa en cada región, así como sobre los efectos relativos de las distintas fuerzas evolutivas que afectan a los niveles de polimorfismo. Por tanto, los estudios de selección a lo largo del genoma deben incorporar el concepto de selección ligada junto con una aproximación que permita detectar y estimar este efecto a lo largo del genoma. Es decir, es necesario conocer aquellas regiones en las que actúa la selección ligada para obtener estimas precisas de los parámetros demográficos y de la acción de la selección natural en el genoma.

#### 4.4. Fiabilidad de los LSB estimados

Los LSB están afectados por los sucesos de recombinación, la tasa de mutación y la intensidad de la selección. El método desarrollado en esta tesis para definir los LSB captura el efecto conjunto de las distintas variables que actúan junto con la recombinación en el genoma. Además, es una aproximación intuitiva y relativamente fácil de aplicar. El método caracteriza los LSB y los NLSB, regiones del genoma que tienen propiedades diferentes y que evolucionan de forma distinta.

Así pues, podemos definir operativamente los LSB como aquellas regiones del genoma que presentan las siguientes características distintivas (i) una correlación entre el polimorfismo y la recombinación estimada en ventanas de observación. (ii) La recombinación corregida por la tasa de mutación (la razón  $c/\mu$ ) es < que 1 y en promedio menor que en los NLSB. Esto indica que en los LSB la importancia relativa de la mutación es superior a la de la recombinación mientras que en los NLSB la fuerza de la recombinación parece ser superior a la de la mutación. (iii) El desequilibrio de ligamiento, medido como  $r^2$ , es mayor que en los NLSB. (iv) No se observa una correlación entre el polimorfismo y la divergencia. (v) La eficiencia de selección tiende a ser menor. En contraposición, en las regiones definidas como NLSB se observan las características inversas a las observadas para los LSB. Todo esto apoya la hipótesis que en los NLSB las ventanas de observación se comportan como si fueran selectivamente independientes

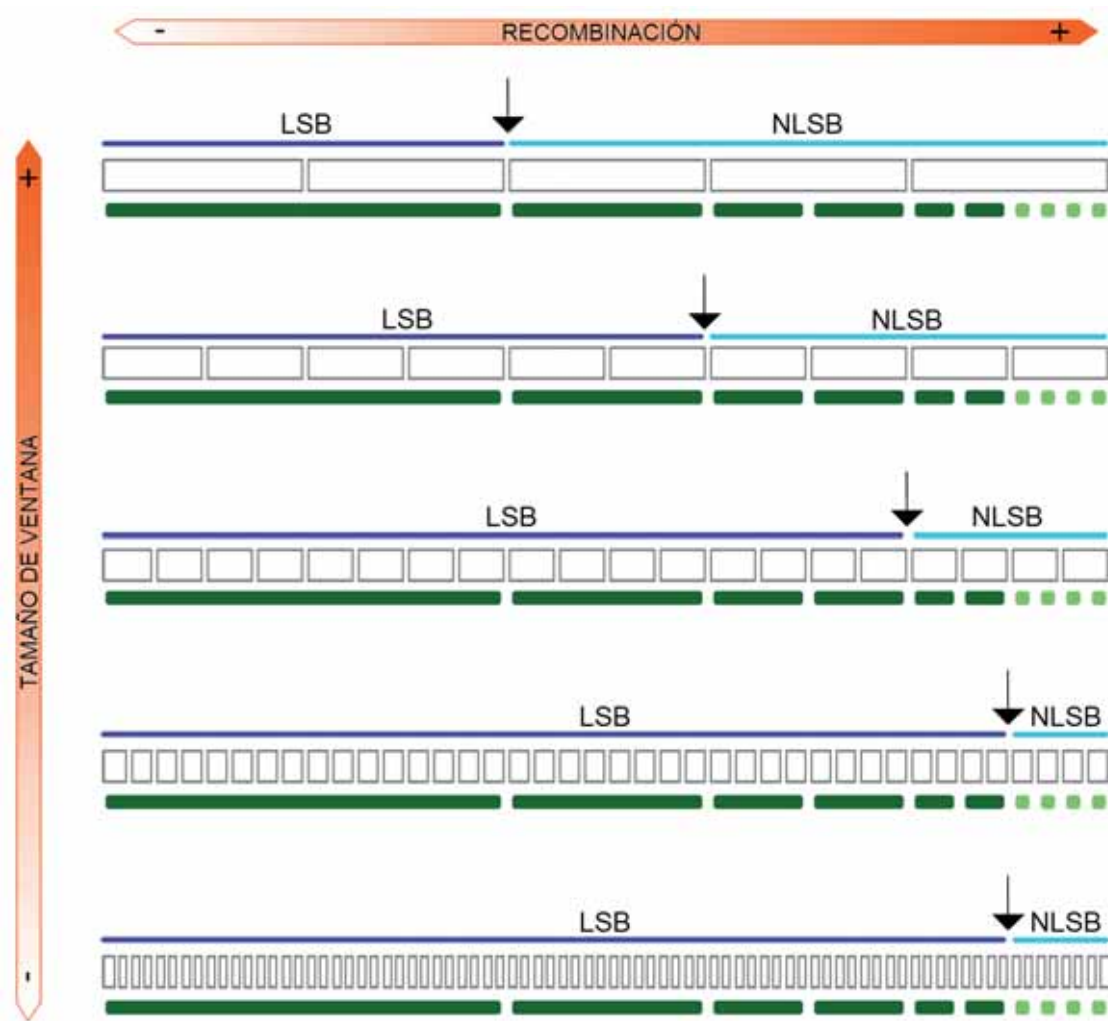
---

unas de otras, haciendo que su evolución sea diferente a los LSB, en los que la selección ligada es la fuerza principal.

Sin embargo, nuestro método presenta las siguientes limitaciones: (i) los resultados son muy dependientes del tamaño de ventana utilizado durante el análisis (sección 4.5). (ii) la independencia del parámetro de recombinación ( $4N_e r$ ) no está garantizada, pues es una medida que puede estar afectada por la selección (Chan *et al.* 2012) (sección 4.6). Además, (iii) las estimas de selección podrían verse afectadas por la selección ligada en las regiones donde la recombinación es nula (Keightley y Eyre-Walker 2007). Aunque teóricamente las estimas de  $\omega_a$  son robustas a las desviaciones introducidas por la selección ligada,  $\alpha$  se subestima cuando la selección ligada es muy fuerte incluso con el método utilizado en este trabajo (Messer y Petrov 2013) (sección 4.7). En las siguientes secciones se tratarán las limitaciones mencionadas con mayor detalle.

#### 4.5. Efecto del tamaño de ventana sobre los LSB

Los LSB se definen como bloques o regiones del genoma donde la selección ligada actúa como fuerza principal. El tamaño de ventana con el que se estima el umbral de recombinación a partir del cual se desacoplan la recombinación y la variación nucleotídica define la longitud de los LSB que hay en una región determinada (sección 3.3.3). Es decir, si en ventanas de 100kb se observa una correlación entre el polimorfismo y la recombinación, esto significa que el efecto de la selección ligada abarca un LSB igual o mayor que 100kb. Por el contrario, la ausencia de correlación entre el polimorfismo y la recombinación en ventanas de 100kb significa que la selección ligada, si ocurre, se da en LSBs cuyo tamaño es menor al tamaño de ventana. El LSB de menor tamaño del genoma lo denominamos como unidad mínima de bloque de selección ligada (minLSB). Teóricamente, esta unidad minLSB debería ser el nucleótido, pero como la recombinación no es infinita o mayor que  $10^8$  (la inversa de la tasa de mutación por nucleótido) la unidad minLSB será superior a 1pb (Figura 4.2). La situación ideal sería poder conocer la unidad mínima de bloque de selección ligada (minLSB) y así ajustar el tamaño de ventana a esa longitud mínima, lo que a su vez permitiría obtener un mapa más preciso de LSB y NLSB. Ahora bien, determinar el minLSB no es trivial. Al ir reduciendo secuencialmente el tamaño de ventana se puede obtener información más precisa del minLSB. De modo que en principio, la mejor forma de estimar la unidad minLSB sería obtener un valor umbral de recombinación que no varíe con las reducciones subsiguientes del tamaño de las ventanas (Figura 4.2). Cuando esto ocurra, el tamaño de la ventana a partir del cuál se fija el valor umbral se correspondería con la unidad minLSB.



**Figura 4.2.** Esquema ilustrativo del efecto del tamaño de ventana sobre la estimación del valor umbral de recombinación

Los distintos LSB de un cromosoma (rectángulos verde oscuro) están ordenados de mayor a menor a lo largo del eje horizontal (de izquierda a derecha). A medida que se reduce el tamaño de ventana de observación (rectángulos blancos), el valor umbral de recombinación estimado (indicado con la flecha) incrementa. Si el tamaño de ventana es igual a la unidad minLSB (rectángulos verde claro), el valor umbral de recombinación estimado es capaz de distinguir todos los LSB. A tamaños de ventana menores no variará el valor umbral de recombinación estimado. Esto se refleja en el gráfico inferior de esta figura.

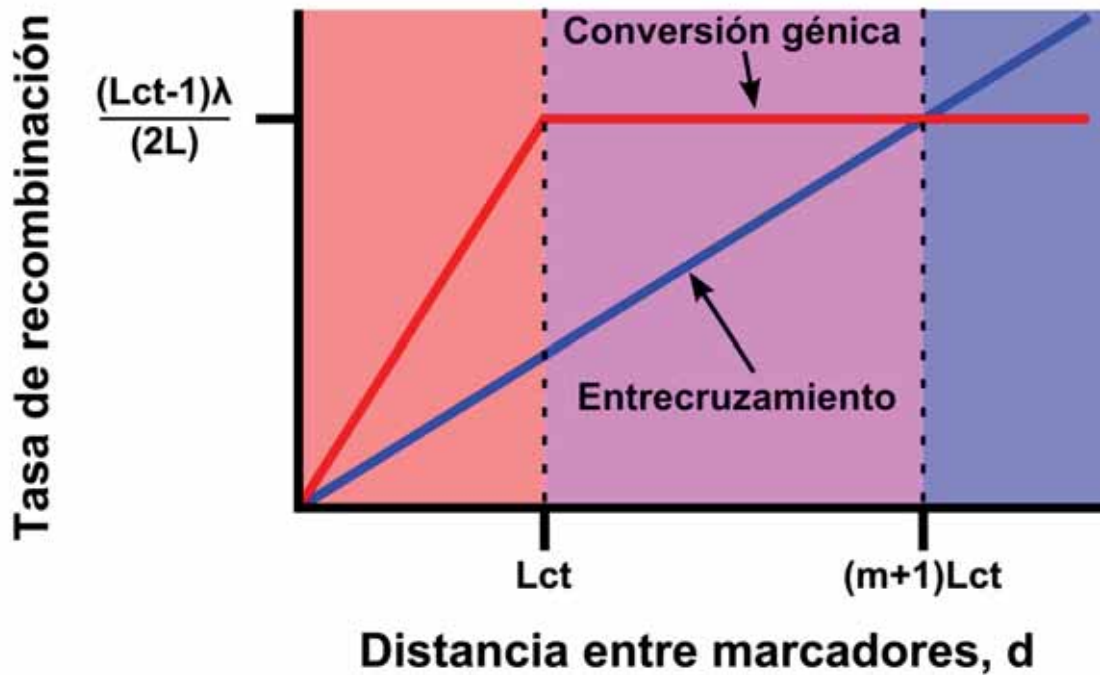
La búsqueda de la unidad minLSB fue lo que nos llevó a estudiar el efecto de las ventanas de observación en el mapa de LSB. Cuando las ventanas son grandes, los resultados pueden ser promedios groseros en los que se pierde información relevante de los LSB menores que el tamaño de ventana. Contrariamente, si las ventanas son muy pequeñas respecto a la longitud del LSB, se incrementa el ruido estadístico, entre otras causas, a que puede haber pocos o ningún sitios polimórficos en las ventanas, reduciendo así la fuerza de la correlación y por ende, la potencia del método de detección de los bloques. En ventanas pequeñas tenemos una mayor resolución, pero debido a una mayor varianza en los datos, la correlación entre estas dos variables es más débil, lo que a su vez significa que estamos explicando una fracción menor de

---

la variación observada. Además, para las ventanas más pequeñas utilizadas en este estudio, las ventanas de 1kb, no pudimos obtener estimas fiables de la tasa de entrecruzamiento y por tanto no pudimos detectar los LSB a este tamaño (sección 3.8).

Para entender que está ocurriendo a resolución de 1kb hay que tener en cuenta la importancia relativa de los dos procesos íntimamente ligados a la recombinación homóloga: la conversión génica y el entrecruzamiento. Cada suceso de recombinación ( $c$ ) puede resolverse como conversión génica asociado a entrecruzamiento o como sucesos de conversión génica sin entrecruzamiento (Mortimer y Fogel 1974, Foss *et al.* 1993, Navarro *et al.* 1996). Berry y Barbadilla (2000) argumentan sobre la importancia relativa de estos dos procesos básicos de recombinación en el nivel de la secuencia de DNA (Figura 4.3). Al considerar las variables longitud promedio de un tracto de conversión ( $Lct$ ) y la razón entre la tasa de conversión génica y la de entrecruzamiento ( $m+1$ ), los autores distinguen tres regiones que caracterizan la importancia relativa de la conversión *vs* el entrecruzamiento como causas de la recombinación entre marcadores: (1) para distancias cortas entre dos marcadores,  $d$  (distancia entre marcadores)  $< Lct$ , la conversión génica es más importante que el entrecruzamiento como fuente de recombinación. (2) A distancias en las que  $d$  es mayor que  $Lct$  pero menor que  $(m+1)Lct$ , la conversión génica sigue generando más recombinación que el entrecruzamiento, aunque el efecto del entrecruzamiento se aproxima al de la conversión linealmente con la distancia. Por último, (3) a distancias largas, mayores que  $(m+1)Lct$ , el entrecruzamiento se convierte en el proceso dominante de recombinación.

La razón de la frecuencia de los sucesos de conversión génica sin entrecruzamiento respecto a la frecuencia de los sucesos que incluyen el entrecruzamiento se ha estimado experimentalmente que es  $\sim 4$  (Hilliker y Chovnick 1981; Hilliker *et al.* 1991; Foss *et al.* 1993; Comeron *et al.* 2012). Es decir, que en *Drosophila* de cada 5 sucesos de conversión génica ocurre un suceso de entrecruzamiento. Además, un estudio reciente (Comeron *et al.* 2012, Cuadro 3) calcula que la longitud promedio del tracto de conversión del genoma es de  $\sim 518$ pb. Teniendo en cuenta estos valores observados, podemos suponer que la conversión génica es 5 veces más importante que el entrecruzamiento en regiones menores que 518pb. Por tanto, la distancia crítica a partir de la cual el entrecruzamiento deja de ser la fuerza principal generadora de recombinación sería:  $518 * 5 = 2590$ pb. Es decir, para distancias menores a  $\sim 2,5$ kb la conversión génica sería la causa principal de recombinación.



**Figura 4.3.** Gráfico que relaciona la distancia entre dos marcadores y su probabilidad que recombinen por conversión génica o por entrecruzamiento.

La línea roja representa la tasa de conversión génica. La línea azul la tasa de entrecruzamiento. El color de fondo delimita las 3 regiones que caracterizan la importancia relativa de la conversión *versus* el entrecruzamiento. (1) Rojo: Gobierna la conversión génica, (2) magenta: El efecto del entrecruzamiento se aproxima al de la conversión y (3) azul: Gobierna el entrecruzamiento.  $L$ : Longitud del cromosoma,  $Lct$ : Longitud del tracto de conversión,  $m$ : Razón de los sucesos de conversión respecto a los sucesos de entrecruzamiento,  $\lambda$ : Tasa de conversión por generación. Figura adaptada de Berry y Barbadilla 2000.

Extrapolando estos resultados a nuestros análisis, en las ventanas de 1kb el proceso de recombinación que debería gobernar es la conversión génica mientras que a partir de las ventanas de 10kb debería ser el entrecruzamiento. Esto estaría de acuerdo con las correlaciones observadas en esta tesis (Tabla 3.25). La correlación entre polimorfismo y recombinación observada en ventanas de 10kb o superior, y que tan repetidamente se ha reportado en la literatura (revisado en Smukowski y Noor 2011; Cutter y Payseur 2013), resulta de la correlación entre el polimorfismo y entrecruzamiento. Sin embargo, la ausencia de correlación en ventanas de 1kb es probablemente debido a que la conversión génica gobierna a distancia menores a 1kb. A distancias  $d$  en el intervalo  $Lct < d < (m+1)Lct$ , la probabilidad de recombinación por conversión es mayor que por entrecruzamiento y la probabilidad es constante, no varía con la distancia (Figura 4.3), por lo que no se espera una correlación entre los patrones de variación y la recombinación a estas pequeñas distancias. La relación podría incluso llegar a ser negativa, pues el número de roturas de doble cadena, es decir el número de sucesos de recombinación en su conjunto, parece ser más o menos constante a lo largo del genoma. Lo que varía es la proporción del proceso resolutivo, la conversión génica o el entrecruzamiento (Comeron *et al.* 2012). Se observa pues que la tasa de conversión génica tiene un ligero



---

incremento en las regiones donde la tasa de entrecruzamiento cae, alrededor de los centrómeros (Comeron *et al.* 2012), lo que podría explicar la correlación negativa observada entre la conversión génica y el polimorfismo. Además, los resultados de las simulaciones llevadas a cabo en esta tesis muestran que para la longitud promedio del tracto de conversión observada en el genoma (~500pb; Comeron *et al.* 2012) se requiere una tasa de conversión génica mucho mayor para ver efectos similares a los del entrecruzamiento sobre el polimorfismo (sección 3.5.1.2, Figura 3.13). Los resultados teóricos y las estimas empíricas obtenidas sugieren que en ventanas de 1kb la fuerza de recombinación que gobierna es la conversión génica y ésta no se correlaciona con los patrones de polimorfismo.

Si la argumentación previa es correcta, entonces hablar de bloques de selección ligada sólo tendría sentido allí donde el entrecruzamiento es la fuerza principal, que según las predicciones se sitúa en regiones cromosómicas mayores de ~2,5kb. Es decir, que el tamaño minLSB se situaría en torno a ~2,5kb. En ese caso, en ventanas de 1kb, como gobierna la conversión génica, la fuerza de la selección ligada debería ser similar en cualquier punto del genoma. Esto es, de hecho, lo que se observa en las estimas del desequilibrio de ligamiento: en ventanas de 1kb no hay diferencias significativas entre los LSB y los NLSB (sección 3.5.2, Figura 3.14, Tabla 3.17).

Supongamos que un suceso de arrastre tiene la misma probabilidad de producirse en cualquier punto del genoma. Su éxito dependerá de las características asociadas a esa mutación como es el coeficiente de selección o su frecuencia inicial en la población (revisado en Orr 2005; Olson-Manning *et al.* 2012). Este suceso de arrastre, ocurra donde ocurra, reducirá los niveles de variación. Ahora bien, dependiendo de la región genómica donde se dé, la magnitud de la reducción será diferente. En los NLSB, esta reducción de la variación no superaría la unidad minLSB. Esta unidad coincidiría con la distancia a la que comienza a gobernar el entrecruzamiento, la estimada ~2,5kb. Por tanto, la reducción neta del polimorfismo en ventanas mayores a la unidad mínima de ligamiento, los supuestos ~2,5kb, sería despreciable. Siguiendo esta lógica, en los NLSB la alta tasa de entrecruzamiento separaría los bloques selectivos mínimos de forma eficiente, impidiendo que los niveles de polimorfismo se redujeran más allá de estos bloques selectivos mínimos.

Sin embargo, en los LSB, la reducción en el polimorfismo provocado por un arrastre superaría la unidad minLSB debido a que la tasa de entrecruzamiento no sería suficientemente frecuente como para independizar regiones genómicas tan pequeñas. Los LSB se alargarían más allá de la unidad minLSB, que si las estimaciones propuestas son correctas, debería ser superior a los 2,5kb, formando bloques más o menos largos en función de la tasa de recombinación. En este contexto, la reducción en el polimorfismo sería importante, puesto que abarcaría una región mayor que la unidad minLSB, dejando huellas claras de la presencia de la selección ligada.

---

Si suponemos que la unidad minLSB es en promedio  $\sim 2,5\text{kb}$ , podemos tratar de ajustar el mapa de BLS y los BSNL a este tamaño de ventana. Directamente no pudimos cartografiar el mapa de  $2,5\text{kb}$ , puesto que el menor tamaño de ventana que pudimos probar es  $10\text{kb}$  (sección 3.8, Tabla 3.25). A esta resolución, se observa que el 20% de los autosomas y 73% de la longitud del cromosoma X están comprendidos por regiones donde la selección ligada es una fuerza secundaria. Ahora bien, nuestros resultados muestran que cuando se disminuye por 10 el tamaño de la ventana de observación los LSB aumentan un 7%. Según esta relación, el pasar de ventanas de  $10\text{kb}$  a  $2,5\text{kb}$  (una reducción de 4 veces el tamaño) supondría un incremento de los LSB de  $\sim 2,8\%$ . Es decir, que en ventanas de  $2,5\text{kb}$  observaríamos que el 82,8% de los autosomas y el 29,8% del cromosoma X estarían formados por LSB. Resumiendo, estimamos que un  $\sim 44\%$  del genoma se encuentra exento de los efectos de la selección ligada. Esta aproximación es más bien conservadora y la proporción LSB podría ser menor, porque en ventanas de  $10\text{kb}$  parece que los LSB contienen NLSB que no hemos podido separar. De cualquier manera, parece que independientemente de la ventana escogida, hay una región del genoma, mayoritaria en el cromosoma X, que no está gobernada por la selección ligada y que por tanto cumple los supuestos predichos por la teoría neutralista (Kimura 1983).

#### 4.6. Relación entre $\rho$ y $c$

Hemos utilizado dos medidas de recombinación diferentes a lo largo del trabajo,  $\rho$  y  $c$ . Mientras que  $\rho$  representa la recombinación histórica debido a que recoge la información a lo largo de los individuos de una población y el tiempo ( $4N_e$  generaciones),  $c$  es una medida basada en la tasa de recombinación actual de ocho individuos (entre 1 y 10 generaciones) (Comeron *et al.* 2012).

En casi todos los análisis realizados, el valor umbral estimado con el parámetro  $\rho$  parece discriminar mejor los LSB frente a los NLSB. Los arrastres por selección positiva pueden crear patrones de desequilibrio de ligamiento muy similar a los producidos por puntos calientes de recombinación (*recombination hotspot*) (Reed y Tishkoff 2006; Stephan *et al.* 2006; Kim y Nielsen 2004), aunque otros autores argumentan que las huellas que dejan los arrastres y los puntos calientes de recombinación son distinguibles (McVean 2007). Chan y colaboradores (2012) realizaron simulaciones para probar la fiabilidad del software *LDhat*. En las simulaciones con selección, las estimas que se obtienen de  $\rho$  son menos precisas que en escenarios sin selección y éstas tienden a estar sobreestimadas cerca de los sitios selectivos. Sin embargo, bajo escenarios de arrastres recurrentes, las estimas de  $\rho$  tienden a subestimarse. El efecto es menor si los arrastres son débiles, aunque sean mucho más frecuentes. En este trabajo no se estudió el efecto de los arrastres por selección negativa. Es decir, que no tenemos una visión completa de cómo puede estar afectando la selección a  $\rho$  en el genoma de *Drosophila melanogaster*. Con los datos disponibles, interpretamos que en los LSB el valor de  $\rho$  podría

---

estar subestimado de forma más severa que en los NLSB, haciendo más evidentes todavía las diferencias entre ellos. Si suponemos que la recombinación es un proceso dinámico que cambia a lo largo del tiempo y entre los individuos de la misma población (Comeron *et al.* 2012) y que además  $\rho$  puede estar afectado por la selección ligada de modo que la recombinación en los LSB se subestima (Chan *et al.* 2012), entonces podemos entender por qué se obtienen resultados más rotundos con  $\rho$ . Por tanto, podemos concluir que para detectar regiones en las que la principal fuerza sea la selección ligada es más apropiado utilizar el parámetro poblacional de recombinación que la tasa de recombinación. Ahora bien, el uso de esta variable dificulta la interpretación de los resultados. Tanto las variaciones en la tasa de recombinación como las variaciones en la intensidad de selección podrían dar lugar a un LSB o un NLSB en el genoma. Nótese que la mutación es otra variable genómica que podría afectar a las estimas de  $\rho$  y por tanto a la detección de los LSB. Sin embargo, no creemos que sea un efecto relevante en este caso puesto que el software con el que se estimó  $\rho$  fue especialmente diseñado para organismos en los que la tasa de mutación es muy elevada, como en virus, bacterias o en los genomas mitocondriales (McVean *et al.* 2002). Además, el polimorfismo está corregido por la divergencia, o lo que es lo mismo, si la divergencia es neutra estamos corrigiendo para la tasa de mutación ( $k = \mu$ ). La otra estima de recombinación,  $c$ , calculada de forma independiente, permite corroborar el efecto principal de la recombinación en la generación de los LSB. Con  $c$  también se observa una buena delimitación de LSB y NLSB, pero estos no están tan claramente definidos en todos los análisis realizados. Esto nos lleva a concluir que, como ya se había visto con el análisis de la varianza a la hora de definir el patrón de diversidad nucleotídica, la recombinación es el factor principal que delimita las regiones afectadas por la selección ligada, aunque la intensidad de la selección o las covariaciones de ambas fuerzas a lo largo del tiempo también afectan la generación y estabilidad de estos bloques a lo largo del genoma.

#### 4.7. Eficiencia de selección

No se sabe hasta qué grado la tasa de recombinación incrementa la eficiencia de selección (Webster y Hurst 2012). La correlación entre la tasa de entrecruzamiento y el polimorfismo a lo largo del genoma puede interpretarse como que la tasa de recombinación regula el grado de interferencia de Hill-Robertson, lo que implica que la eficiencia de la selección varíe a lo largo del genoma. Sin embargo, otros autores consideran que observar dicha correlación no lleva necesariamente a concluir que la variación de la eficiencia de selección se deba a la tasa de recombinación (Webster y Hurst 2012). Es bien conocido que las regiones de los genomas que no recombinan acumulan alelos deletéreos y degeneran con el tiempo (Charlesworth y Charlesworth 2000; Bachtrog y Charlesworth 2002; Bachtrog 2005; Bartolomé y Charlesworth 2006; Haddrill *et al.* 2007). Además, se ha descrito un incremento de la razón  $k_a/k_s$  en estas regiones, pero no se ha encontrado una variación significativa en el resto del genoma (McVean

---

y Charlesworth 2000; Bachtrog 2003; Haddrill *et al.* 2007). Esta observación apoyaría la teoría de que pequeños incrementos (en torno a 0,5 cM/Mb) en la tasa de recombinación son suficientes para experimentar la mayoría de los beneficios asociados a la recombinación (Hurst y Peck 1996). Sin embargo, otros autores argumentan que la recombinación incrementa la eficacia de la selección natural para diferentes valores de recombinación (Betancourt y Presgrave 2002; Presgrave 2005; Zhang y Li 2004; Larracunte 2008). Aunque ha habido cierto debate para determinar la importancia relativa de la recombinación en aumentar la eficiencia de la selección (Webster y Husrt 2012), en un trabajo reciente y exhaustivo en el nivel genómico (Campos *et al.* 2014), se ha observado que la recombinación incrementa significativamente la eficiencia de la selección, no sólo al comparar ausencia/presencia de la recombinación, sino también para distintos intervalos de valores de la tasa de recombinación.

En este trabajo se superan ciertas limitaciones de los estudios previos debido a que: (1) se usan datos genómicos que permiten separar las mutaciones ligeramente deletéreas de las adaptativas que se encuentran confundidas en la razón  $k_a/k_s$ , usando información tanto del polimorfismo como de la divergencia. (2) Se utiliza la tasa de recombinación actual e histórica. Ambas contienen información distinta y permiten inferir si la tasa de recombinación es lábil evolutivamente (Pink *et al.* 2009). (3) El mapa de LSB nos permite distinguir categorías o *bins* de recombinación específicos de la población y brazo cromosómico en vez de utilizar categorías arbitrarias para definir regiones de recombinación elevadas *versus* reducidas en el genoma (ej. Larracunte *et al.* 2008) (4) Los métodos basados en MKT pueden dar estimas sesgadas cuando la selección ligada es una fuerza importante. El método empleado en esta tesis parece predecir de forma correcta el efecto de la selección positiva incluso bajo condiciones de selección ligada (Messer y Petrov 2013).

Como se esperaba, en los NLSB se observa una mayor eficiencia de selección. A pesar de que las estimas de  $k_a/k_s$  prácticamente no varían en ambos bloques, la contribución relativa de alelos neutros y/o ligeramente deletéreos y de alelos adaptativos es diferente. En las regiones donde la selección ligada actúa como fuerza principal, la aportación de los alelos neutros o ligeramente deletéreos a la evolución de estas regiones es mayor (~60%) que la de los alelos adaptativos (~40%). Mientras que en las regiones donde la selección ligada es insignificante, la contribución de los alelos adaptativos a los cambios divergentes globales es mayor (~62%) que la de los alelos ligeramente deletéreos o neutros (~38%). Por lo tanto, queda patente la relevancia de utilizar las medidas de selección que distinguen entre la contribución de los alelos ligeramente deletéreos y los adaptativos, pues ambas pueden incrementar  $k_a/k_s$  con significados biológicos opuestos. Estas diferencias se explican principalmente por las grandes porciones centroméricas donde la recombinación se encuentra especialmente reducida. Cuando se eliminan estas regiones (un 30% del genoma), se sigue observando la tendencia general que en

---

los NLSB la eficiencia de selección es mayor, aunque las diferencias son menores y no siempre significativas. Por lo tanto, nuestros datos parecen apoyar que el incremento de la eficiencia de la selección no aumenta de forma lineal o proporcional con la recombinación. Los efectos más dramáticos se observan cuando la recombinación es realmente baja y le sigue un incremento discreto de la eficiencia de selección hasta alcanzar su máximo: aquel valor en el que la recombinación es suficientemente elevada para eliminar los efectos de la iHR.

Estos análisis nos han permitido también observar los cambios tan bruscos que puede sufrir las medidas de  $\alpha$  en función del tamaño de ventana. Esto podría contribuir a explicar las discrepancias encontradas entre diversos estudios de la misma población (Tabla 1.4). Por ejemplo, analizando la misma población norteamericana utilizada en esta tesis con una aproximación diferente (sección 1.2.3), Mackay *et al.* (2012) estiman un valor de  $\alpha \sim 24\%$  en las regiones codificadoras basadas en ventanas de 50kb. Este valor es intermedio que los estimados en este trabajo a 100kb ( $\alpha = \sim 40\%$ ) y a 10kb ( $\alpha = \sim 10\%$ ). Parece, por tanto, que las mayores diferencias provienen del tamaño de ventana utilizado y no del método utilizado para inferir  $\alpha$ . A diferencia de otros estimadores cuyo rango de valores están delimitados de 0 a 1 (como  $\omega_a$  o  $\omega_d$ ) el rango de valores que puede tomar  $\alpha$  es mucho más amplio (intervalo  $(-\infty, 1]$ ). En ventanas pequeñas las estimas de  $\alpha$  pueden tener mayor varianza y los valores extremos de  $\alpha$  puede ser tan negativos que hagan que la media sea más reducida. Hay que tener en cuenta que la longitud promedio de los genes en *Drosophila* es de  $\sim 11,3$ kb (Lewin 1994), por lo tanto, si la selección ocurriera a nivel génico, las ventanas de 10kb deberían ser las más adecuadas.

$\omega_a$  y  $\omega_d$ , no varían en función de la ventana de observación, aunque las estimas de  $\omega_a$  podrían estar afectadas por la selección ligada y por tanto subestimadas cuando la recombinación fuera nula (Keightley y Eyre-Walker 2007). Si esto es así, las diferencias observadas en  $\omega_a$  entre LSB y NLSB serán mayores a las que realmente son, en tanto que las diferencias entre LSB y NLSB para  $\omega_d$  serán menores a las que realmente son. En otras palabras, la disminución de las fijaciones por selección adaptativa en los LSB no sería tan marcada como la descrita en este trabajo. Sin embargo, las diferencias en relación a los alelos ligeramente deletéreos serían mayores que las observadas. Cuando quitamos los centrómeros del análisis, se quitan la mayoría de las regiones donde no hay recombinación y por tanto donde puede estar introduciéndose el sesgo en la estimación de  $\omega_a$  y  $\omega_d$ . En ambos casos las diferencias entre LSB y NLSB disminuyen y por tanto, el efecto de la selección ligada sobre los bloques y su consecuente disminución de la eficiencia de selección va más allá del sesgo que puedan estar introduciendo las estimas. Además, las estimas de  $\omega_d$  disminuyen ligeramente más que las estimas de  $\omega_a$  en las comparaciones con y sin centrómeros (diferencias con y sin centrómeros;  $\omega_d$ : 2,7 veces menor;  $\omega_a$ : 2,1 veces menor), lo contrario de lo que esperaríamos si el sesgo

---

introducido en las estimas de la eficiencia de selección en LSB fuera importante. En consecuencia, no creemos que este factor sea relevante ni una limitación a este trabajo.

#### 4.8. Singularidad del cromosoma X

Los patrones de polimorfismo y divergencia observados en el X son muy diferentes al resto de los autosomas. Observamos que respecto a los autosomas, todo polimorfismo excepto el sinónimo está reducido en este cromosoma mientras que la divergencia (exceptuando también la sinónima) es mayor. Esta observación a escala de todo el genoma permite aclarar los resultados contradictorios obtenidos en diversos estudios previos (Thornton *et al.* 2006; Connallon 2007; Baines *et al.* 2008) en relación a la hipótesis de la evolución rápida del X (Charlesworth *et al.* 1987). Tras analizar el genoma completo de *D. melanogaster* de decenas de individuos, todo parece indicar que el cromosoma X evoluciona globalmente de forma más rápida que los autosomas.

El resto de las observaciones realizadas también apoyan esta hipótesis de la evolución rápida del X. Destaca como principal diferencia respecto a los autosomas que no se observa una clara reducción del polimorfismo en la región centromérica. Esta diferencia no se puede explicar por cambios en la tasa de mutación, debido a que no encontramos diferencias basándonos en la divergencia de los sitios 4 veces degenerados y tampoco se han reportado diferencias entre autosomas y el cromosoma X en experimentos de acumulación de mutaciones (Keightley *et al.* 2009; Schrider *et al.* 2013). Aunque a primera vista no lo parezca, la variable genómica clave que parece explicar esta observación es nuevamente la recombinación (sección 3.2).

El patrón de la recombinación del X es similar al de los autosomas, con un descenso paulatino hacia el centrómero y un descenso brusco en los telómeros (Figura 3.8) y es, como se ha visto, una de las principales variables que explica el patrón de polimorfismo, incluyendo al cromosoma X, aunque con un peso relativo menor (Tabla 3.5). La diferencia entre los autosomas y el X radica en que la tasa de recombinación y el parámetro de recombinación observados en el cromosoma X son 1,8 y 1,5 veces superior al de los autosomas, respectivamente. Esto parece generar una reducción muy substancial del efecto de la selección ligada, que afecta a un ~50% menos de los sitios comparado con los autosomas. Es decir, que a pesar de que el patrón de recombinación es similar, la mayor tasa de entrecruzamiento observada en el cromosoma X resulta en un incremento de ~50% de sitios selectivamente independientes en este cromosoma respecto a los autosomas. Esta podría ser la razón por la que no se observa una disminución del polimorfismo en el tercio centromérico del cromosoma X. Nótese que otras variables, tales como una reducción de los alelos ligeramente deletéreos en relación a los autosomas debido a la condición homocigótica del cromosoma X (Charlesworth *et*

---

al. 1987, Campos *et al.* 2014), podrían también generar este patrón de reducción de sitios segregantes ligados.

El hecho de que la selección ligada afecte a menos sitios en este cromosoma, a su vez, podría generar un aumento de la eficiencia de selección global del cromosoma X debido a una reducción en la iHR. Es decir, que esta observación podría en parte explicar la rápida evolución del cromosoma X. El incremento en la eficiencia de selección en los NLSB respecto a los LSB observados en los autosomas es proporcional al observado en el cromosoma X respecto a los autosomas. Sin embargo, la longitud promedio de los LSB (~364kb menor) y la proporción de sitios que ocupan son mucho menores (~50% como se ha discutido en el párrafo anterior). Parece por tanto, que una mayor recombinación podría reducir el efecto global de la selección ligada, reduciendo la longitud y la proporción de sitios afectados por la selección ligada, e incrementar la eficiencia de la selección en el cromosoma completo. Esto haría que la eficiencia de selección fuera mayor en los LSB y en los NLSB, pero la diferencia entre ambos bloques fuera equivalente al observado en los autosomas. Si esto es así, esperaríamos que los indicadores de la eficiencia de la selección sean superiores en este cromosoma; tanto para la selección positiva como la negativa. Estudios previos de genómica de poblaciones en *Drosophila* apuntan a una mayor eficiencia de la selección natural o incremento de la selección positiva en el cromosoma X (Mackay *et al.* 2012, Langley *et al.* 2012, Pool *et al.* 2012). Sin embargo, en esta tesis se observa que los valores de sustituciones neutras y/o ligeramente deletéreas son similares entre autosomas y X ( $A\omega_d$ : 0,0916 y  $X\omega_d$ : 0,0912). Hay diferencias significativas para las fijaciones de alelos adaptativos entre autosomas y X ( $A\omega_a$ : 0,0894 y  $X\omega_a$ : 0,1190;  $A\alpha$ : 0,3773  $X\alpha$ : 0,5116), pero al mirar cromosoma a cromosoma estos valores tampoco son diferentes al mayor de los autosomas; el cromosoma 2L ( $2L\omega_a$ : 0,1093 y  $X\omega_a$ : 0,1190;  $2L\alpha$ : 0,4610  $X\alpha$ : 0,5116, Tabla suplementaria 3).

Teóricamente, el hecho de que el cromosoma X sea hemicígote en machos hace que su tamaño efectivo sea  $\frac{3}{4}$  de la de los autosomas. Esta reducción del censo efectivo supondría una disminución en el polimorfismo de este cromosoma y también en la eficiencia de selección. De hecho, el cromosoma 3R que muestra un patrón de polimorfismo similar al X corregido, aunque tiene una mayor cantidad de sustituciones ligeramente deletéreas y una de las menores fijaciones de alelos adaptativos respecto a las sustituciones neutras (Tabla suplementaria 3). Es decir, que el 3R es el autosoma con la menor eficiencia de selección. Sin embargo, el cromosoma X, a pesar de tener un polimorfismo equiparable al menos polimórfico de los autosomas (tras la corrección en el censo efectivo) tiene una eficiencia de selección equiparable al de mayor eficiencia de los autosomas (el 2L). Teniendo todo esto en cuenta, la eficiencia de selección en el cromosoma X es superior a la de los autosomas, a pesar de su reducido tamaño de población, suponemos que por la selección ligada disminuida que presenta.

---

Es decir, que el cromosoma X a pesar de tener un censo efectivo menor al de los autosomas y un polimorfismo reducido, exhibe una tasa de evolución molecular y una eficiencia de la selección mayor que los autosomas, lo que apoya la hipótesis de la rápida evolución del X (Charlesworth *et al.* 1987). Las causas de esta rápida evolución del X respecto a los autosomas parece ser la menor intensidad de la selección ligada observada en este cromosoma gracias en parte a una mayor tasa de recombinación. Por otro lado, existe una hipótesis alternativa pero no excluyente que podría explicar el posible incremento en la eficiencia de selección del cromosoma X. Se ha argumentado que las nuevas mutaciones ligadas al X al estar directamente expuestas a la selección en los machos hemicígotos, incrementando de este modo la eficiencia de selección en el cromosoma X (Charlesworth *et al.* 1987).



---

## Parte 5

# Conclusiones



---

## 5. Conclusiones

- i) Se ha llevado a cabo un estudio exhaustivo de genómica de poblaciones en el que se describen los patrones de variación genómico de los genes codificadores de *Drosophila melanogaster*. En el tercio centromérico de cada brazo autosómico, el polimorfismo está reducido en un ~53% y la divergencia aumenta ligeramente, ~10%. Los patrones de  $k_0/k_4$  y  $\pi_0/\pi_4$  parecen indicar una reducción drástica de la eficiencia de la selección en esta región.
- ii) La tasa de recombinación es la principal variable explicativa de los patrones de variación nucleotídica en el genoma de *Drosophila melanogaster*, tanto en los cromosomas autosómicos como en el cromosoma X, con independencia del patrón distinto de polimorfismo observado en el cromosoma X.
- iii) La correlación universalmente hallada entre el polimorfismo y la recombinación desaparece a partir de un valor umbral de recombinación en el nivel del cromosoma. Este valor umbral se ha estimado en los diferentes brazos cromosómicos y su valor varía entre 3,36 cM/Mb del brazo 2R a 1,54 cM/Mb del cromosoma X.
- iv) Dos nuevos conceptos de genómica de poblaciones se aportan en esta tesis: Bloque de selección Ligada (LSB) y Bloque sin Selección Ligada (NLSB). El genoma de *D. melanogaster* es un mosaico constituido por ambos tipos de bloques en los que la dinámica de la evolución molecular es opuesta y en consecuencia deben ser reconocidos y cartografiados.
- v) Los LSB abarcan la mayor parte de los autosomas, un ~73%, y una fracción menor del cromosoma X, ~23%. Los NLSB, que representan un 40% del genoma, son regiones del genoma donde no hay evidencia de selección ligada al menos a los tamaños de las ventanas de observación con los que se analizaron los datos.
- vi) Los LSB y NLSB presentan las siguientes características distintivas: i) la razón  $c/\mu$  es en promedio menor a 1 en los LSB y superior a 1 en los NLSB. ii) El desequilibrio de ligamiento es mayor en los LSB y menor en los NLSB. iii) En los NLSB, el polimorfismo y la divergencia están principalmente afectados por la tasa de mutación, generando una correlación positiva entre ambas que no se observa en los LSB debido a la selección ligada. iv) La eficiencia de selección es mayor en los NLSB que en los LSB y esta tendencia se observa más allá de los enormes bloques centroméricos.

- 
- vii) Se propone la existencia de una unidad mínima de bloque de selección ligada a lo largo del genoma. Esta unidad minLSB vendría determinada por la importancia relativa de la conversión génica y el entrecruzamiento. A distancias en las que gobierna la conversión génica (que hemos estimado de alrededor 2,5kb) el nivel de ligamiento será el mínimo y constante a lo largo del genoma. A distancias en las que el entrecruzamiento es la principal fuerza de recombinación, o bien las distintas unidades minLSB se mantendrán ligadas formando grandes LSB o bien allí donde la tasa de entrecruzamiento sea elevada las unidades minLSB se mantendrán selectivamente independientes (NLSB).
- viii) El cromosoma X exhibe unos patrones de variación y LSB singulares. Nuestros datos corroboran la hipótesis de la evolución rápida del X. En el cromosoma X hay una reducción del ~50% en el número de sitios afectados por la selección ligada respecto a los autosomas. A pesar de que el patrón de recombinación es similar al de los autosomas, la mayor tasa de entrecruzamiento observada en el cromosoma X podría explicar este incremento en el tamaño de los NLSB. El reducido número de LSB podría ser una de las causas de la evolución rápida del X.

---

# Bibliografía



---

## 6. Bibliografía

- Adams, M. D., S. E. Celniker, R. A. Holt, C. A. Evans, J. D. Gocayne, P. G. Amanatides, S. E. Scherer, *et al.* 2000. The Genome Sequence of *Drosophila Melanogaster*. *Science* **287** (5461): 2185–95.
- Aguade, M., N. Miyashita, y C. H. Langley. 1989. Reduced Variation in the Yellow-Achaete-Scute Region in Natural Populations of *Drosophila Melanogaster*. *Genetics* **122** (3): 607–615.
- Akaike, H. 1974. A New Look at the Statistical Model Identification. *IEEE Transactions on Automatic Control* **19** (6): 716–723.
- Akashi, H. 1995. Inferring Weak Selection from Patterns of Polymorphism and Divergence at ‘Silent’ Sites in *Drosophila* DNA. *Genetics* **139** (2): 1067–1076.
- Akashi, H. 1996. Molecular Evolution between *Drosophila Melanogaster* and *D. Simulans*: Reduced Codon Bias, Faster Rates of Amino Acid Substitution, and Larger Proteins in *D. Melanogaster*. *Genetics* **144** (3): 1297–1307.
- Akashi, H., y A. Eyre-Walker. 1998. Translational Selection and Molecular Evolution. *Current Opinion in Genetics & Development* **8** (6): 688–693.
- Akey, J. M., R. Nielsen, M. J. Hubisz, I. Hellmann, J. K. Pickrell, G. Coop, J. Novembre, S. Mallick, S. Gnerre, y P. Muller. 2009. Constructing Genomic Maps of Positive Selection in Humans : Where Do We Go from Here? *Genome research* **19** (206): 711–722.
- Akey, J. M., M. A. Eberle, M. J. Rieder, C. S. Carlson, M. D. Shriver, D. A. Nickerson, y L. Kruglyak. 2004. Population History and Natural Selection Shape Patterns of Genetic Variation in 132 Genes. *PLoS Biology* **2** (10): e286. doi: 10.1371/journal.pbio.0020286
- Alves, J. M., L. Chikhi, A. Amorim, y A. M. Lopes. 2014. The 8p23 Inversion Polymorphism Determines Local Recombination Heterogeneity across Human Populations. *Genome Biology and Evolution* **6** (4): 921–930.
- Andolfatto, P., y M. Nordborg. 1998. The Effect of Gene Conversion on Intralocus Associations. *Genetics* **148** (2): 1397–1399.
- Andolfatto, P. 2001. Contrasting Patterns of X-Linked and Autosomal Nucleotide Variation in *Drosophila Melanogaster* and *Drosophila Simulans*. *Molecular Biology and Evolution* **18** (3): 279–290.
- Andolfatto, P. 2005. Adaptive Evolution of Non-Coding DNA in *Drosophila*. *Nature* **437** (7062): 1149–1152.
- Andolfatto, P. 2007. Hitchhiking Effects of Recurrent Beneficial Amino Acid Substitutions in the *Drosophila Melanogaster* Genome. *Genome Research* **17** (858): 1755–1762.
- Andolfatto, P., K. M. Wong, y D. Bachtrog. 2011. Effective Population Size and the Efficacy of Selection on the X Chromosomes of Two Closely Related *Drosophila* Species. *Genome Biology and Evolution* **3** (1): 114–128. doi:10.1093/gbe/evq086.
- Ashburner M., K. G. Golic, y R. S. Hawley. 2005. *Drosophila: A Laboratory Handbook*. Cole Spring Harbour Laboratory Press, New York.
- Ashburner, M., y C. M. Bergman. 2005. *Drosophila Melanogaster*: A Case Study of a Model Genomic Sequence and Its Consequences. *Genome Research* **15** (12): 1661–1667. doi:10.1101/gr.3726705.15.
- Axelsson, E., M. T. Webster, A. Ratnakumar, C. P. Ponting, y K. Lindblad-Toh. 2012. Death of PRDM9 Coincides with Stabilization of the Recombination Landscape in the Dog Genome. *Genome Research* **22** (1): 51–63. doi:10.1101/gr.124123.111.
- Ayala, F. J., M. L. Tracey, L. G. Barr, J. F. McDonald, y S. Pérez-Salas. 1974. Genetic Variation in Natural Populations of Five *Drosophila* Species and the Hypothesis of the Selective Neutrality of Protein Polymorphisms. *Genetics* **77** (2): 343–84.
- Bachtrog, D., y B. Charlesworth. 2002. Reduced Adaptation of a Non-Recombining Neo-Y Chromosome. *Nature* **416** (6878): 323–326. doi:10.1038/416323a.

- 
- Bachtrog, D. 2003. Adaptation Shapes Patterns of Genome Evolution on Sexual and Asexual Chromosomes in *Drosophila*. *Nature Genetics* **34** (2): 215–219. doi:10.1038/ng1164.
- Bachtrog, D. 2005. Sex Chromosome Evolution: Molecular Aspects of Y-Chromosome Degeneration in *Drosophila*. *Genome Research* **15** (10): 1393–1401. doi:10.1101/gr.3543605.
- Baines, J. F., S. A. Sawyer, D. L. Hartl, y J. Parsch. 2008. Effects of X-Linkage and Sex-Biased Gene Expression on the Rate of Adaptive Protein Evolution in *Drosophila*. *Molecular Biology and Evolution* **25** (8): 1639–1650. doi:10.1093/molbev/msn111.
- Barnes, T. M., Y. Kohara, A. Coulson, y S. Hekimi. 1995. Meiotic Recombination, Noncoding DNA and Genomic Organization in *Caenorhabditis Elegans*. *Genetics* **141** (1): 159–179.
- Bartolomé, C., y B. Charlesworth. 2006. Evolution of Amino-Acid Sequences and Codon Usage on the *Drosophila* Miranda Neo-Sex Chromosomes. *Genetics* **174** (4): 2033–2044. doi:10.1534/genetics.106.064113.
- Baudat, F., y S. Keeney. 2001. Meiotic Recombination: Making and Breaking Go Hand in Hand. *Current Biology* **11** (2): R45–R48. doi:10.1016/S0960-9822(01)00013-6.
- Baudry, E., B. Viginier, y M. Veuille. 2004. Non-African Populations of *Drosophila Melanogaster* Have a Unique Origin. *Molecular Biology and Evolution* **21** (8): 1482–1491. doi:10.1093/molbev/msh089.
- Bauer, V. L., y C. F. Aquadro. 1997. Rates of DNA Sequence Evolution Are Not Sex-Biased in *Drosophila Melanogaster* and *D. Simulans*. *Molecular Biology and Evolution* **14** (12): 1252–1257. doi:10.1093/oxfordjournals.molbev.a025734.
- Bazin, E., S. Glémin, y N. Galtier. 2006. Population Size Does Not Influence Mitochondrial Genetic Diversity in Animals. *Science* **312** (5773): 570.
- Bedada, G., A. Westerbergh, T. Müller, E. Galkin, E. Bdolach, M. Moshelion, E. Fridman, y K. J. Schmid. 2014. Transcriptome Sequencing of Two Wild Barley (*Hordeum Spontaneum* L.) Ecotypes Differentially Adapted to Drought Stress Reveals Ecotype-Specific Transcripts. *BMC Genomics* **15** (1): 995.
- Begun, D. J., y C. F. Aquadro. 1992. Levels of Naturally Occurring DNA Polymorphism Correlate with Recombination Rates in *D. Melanogaster*. *Nature* **356** (6369): 519–520.
- Begun, D. J., y C. F. Aquadro. 1993. African and North American Populations of *Drosophila Melanogaster* Are Very Different at the DNA Level. *Nature* **365**: 548–550.
- Begun, D. J., y P. Whitley. 2000. Reduced X-Linked Nucleotide Polymorphism in *Drosophila Simulans*. *Proc Natl Acad Sci USA* **97** (11): 5960–5965. doi:10.1073/pnas.97.11.5960.
- Begun, D. J., A. K. Holloway, K. Stevens, L. W. Hillier, Y.P. Poh, M. W. Hahn, P. M. Nista, *et al.* 2007. Population Genomics: Whole-Genome Analysis of Polymorphism and Divergence in *Drosophila Simulans*. *PLoS Biology* **5** (11): e310. doi:10.1371/journal.pbio.0050310.
- Berger, J., T. Suzuki, K. A. Senti, J. Stubbs, G. Schaffner, y B. J. Dickson. 2001. Genetic Mapping with SNP Markers in *Drosophila*. *Nature Genetics* **29** (4): 475–481. doi:10.1038/ng773.
- Bergman, C. M., Joseph W. C., y S. E. Celniker. 2005. *Drosophila* DNase I Footprint Database: A Systematic Genome Annotation of Transcription Factor Binding Sites in the Fruitfly, *Drosophila Melanogaster*. *Bioinformatics* **21** (8): 1747–1749. doi:10.1093/bioinformatics/bti173.
- Bernardi, G. 2007. The Neoselectionist Theory of Genome Evolution. *Proc Natl Acad Sci USA* **104** (20) (May 15): 8385–90. doi:10.1073/pnas.0701652104.
- Berry, A. J., J. W. Ajioka, y M. Kreitman. 1991. Lack of Polymorphism on the *Drosophila* Fourth Chromosome Resulting from Selection. *Genetics* **129** (4): 1111–1117.
- Berry, A., y A. Barbadilla. 2000. Gene Conversion Is a Major Determinant of Genetic Diversity at the DNA Level. pág. 102–123 en *Evolutionary Genetics: From Molecules to Morphology*, editado por by R.S. Singh y C.B. Krimbas. Cambridge University Press, USA.
- Betancourt, A. J., y D. C. Presgraves. 2002. Linkage Limits the Power of Natural Selection in *Drosophila*. *Proc Natl Acad Sci USA* **99** (21): 13616–20. doi:10.1073/pnas.212277199.



- 
- Betancourt, A. J., Yuseob K., y H. A. Orr. 2004. A Pseudohitchhiking Model of X vs. Autosomal Diversity. *Genetics* **168** (4): 2261–2269. doi:10.1534/genetics.104.030999.
- Betancourt, A. J., J. J. Welch, y B. Charlesworth. 2009. Reduced Effectiveness of Selection Caused by a Lack of Recombination. *Current Biology* **19** (8): 655–660. doi:10.1016/j.cub.2009.02.039.
- Betrán, E., J. Rozas, A. Navarro, y A. Barbadilla. 1997. The Estimation of the Number and the Length Distribution of Gene Conversion Tracts from Population DNA Sequence Data. *Genetics* **146** (1): 89–99.
- Birky, C. W., y J. B. Walsh. 1988. Effects of Linkage on Rates of Molecular Evolution. *Proc Natl Acad Sci USA* **85** (17): 6414–6418. doi:10.1073/pnas.85.17.6414.
- Biswas, S., y J. M. Akey. 2006. Genomic Insights into Positive Selection. *Trends in Genetics* **22** (8): 437–446. doi:10.1016/j.tig.2006.06.005.
- Blanton, H. L., S. J. Radford, S. McMahan, H. M. Kearney, J. G. Ibrahim, y J. Sekelsky. 2005. REC, Drosophila MCM8, Drives Formation of Meiotic Crossovers. *PLoS Genetics* **1** (3): 343–354. doi:10.1371/journal.pgen.0010040.
- Boyko, A. R., S. H. Williamson, A. R. Indap, J. D. Degenhardt, R. D. Hernandez, K. E. Lohmueller, M. D. Adams, *et al.* 2008. Assessing the Evolutionary Impact of Amino Acid Mutations in the Human Genome. *PLoS Genetics* **4** (5): e100083. doi:10.1371/journal.pgen.1000083.
- Bromham, L., y D. Penny. 2003. The Modern Molecular Clock. *Nature Reviews. Genetics* **4** (3): 216–224. doi:10.1038/nrg1020.
- Brunschwig, H., L. Levi, E. Ben-David, R. W. Williams, B. Yakir, y S. Shifman. 2012. Fine-Scale Maps of Recombination Rates and Hotspots in the Mouse Genome. *Genetics* **191** (3): 757–764. doi:10.1534/genetics.112.141036.
- Bussell, J. J., N. M. Pearson, R. Kanda, D. A. Filatov, y B. T. Lahn. 2006. Human Polymorphism and Human-Chimpanzee Divergence in Pseudoautosomal Region Correlate with Local Recombination Rate. *Gene* **368** (1-2): 94–100. doi:10.1016/j.gene.2005.10.020.
- Bustamante, C. D., A. Fledel-Alon, S. Williamson, R. Nielsen, M. T. Hubisz, S. Glanowski, D. M. Tanenbaum, *et al.* 2005. Natural Selection on Protein-Coding Genes in the Human Genome. *Nature* **437** (7062): 1153–1157. doi:10.1038/nature04240.
- Bustamante, C. D., R. Nielsen, S. A. Sawyer, K. M. Olsen, M. D. Purugganan, y D. L. Hartl. 2002. The Cost of Inbreeding in Arabidopsis. *Nature* **416** (6880): 531–534. doi:10.1038/416531a.
- Bustamante, C. D., J. Wakeley, S. Sawyer, y D. L. Hartl. 2001. Directional Selection and the Site-Frequency Spectrum. *Genetics* **159** (4): 1779–1788.
- Caballero, A. 1995. On the Effective Size of Populations with Separate Sexes, with Particular Reference to Sex-Linked Genes. *Genetics* **139** (2): 1007–1011.
- Campos, J. L., D. L. Halligan, P. R. Haddrill, y B. Charlesworth. 2014. The Relation between Recombination Rate and Patterns of Molecular Evolution and Variation in *Drosophila Melanogaster*. *Molecular Biology and Evolution* **31** (4): 1010–1028. doi:10.1093/molbev/msu056.
- Canty, A., y B. Ripley. 2015. Boot: Bootstrap R (S-Plus) Functions. R package version 1.3-16.
- Caracristi, G., y C. Schlötterer. 2003. Genetic Differentiation between American and European *Drosophila Melanogaster* Populations Could Be Attributed to Admixture of African Alleles. *Molecular Biology and Evolution* **20** (5): 792–799. doi:10.1093/molbev/msh091.
- Carneiro, M., F. W. Albert, J. Melo-Ferreira, N. Galtier, P. Gayral, J. A. Blanco-Aguilar, R. Villafuerte, M. W. Nachman, y N. Ferrand. 2012. Evidence for Widespread Positive and Purifying Selection across the European Rabbit (*oryctolagus Cuniculus*) Genome. *Molecular Biology and Evolution* **29** (7): 1837–1849. doi:10.1093/molbev/mss025.

- 
- Carvalho, A. B., M. D. Vibranovski, J. W. Carlson, S. E. Celniker, R. A. Hoskins, G. M. Rubin, G. G. Sutton, M. D. Adams, E. W. Myers, y A. G. Clark. 2003. Y Chromosome and Other Heterochromatic Sequences of the *Drosophila Melanogaster* Genome: How Far Can We Go? *Genetica* **117** (2-3): 227–237. doi:10.1023/A:1022900313650.
- Casillas, S., y A. Barbadilla. 2006. PDA v.2: Improving the Exploration and Estimation of Nucleotide Polymorphism in Large Datasets of Heterogeneous DNA. *Nucleic Acids Research* **34**: 632–634. doi:10.1093/nar/gkl080.
- Casillas, S., A. Barbadilla, y C. M. Bergman. 2007. Purifying Selection Maintains Highly Conserved Noncoding Sequences in *Drosophila*. *Molecular Biology and Evolution* **24** (10): 2222–2234. doi:10.1093/molbev/msm150.
- Cavalli-Sforza, L. L. 1966. Population Structure and Human Evolution. *Proc. R. Soc. Lond. B* **164** (995): 362–379.
- Celniker, S. E., D. A. Wheeler, B. Kronmiller, J. W. Carlson, A. Halpern, S. Patel, M. Adams, *et al.* 2002. Finishing a Whole-Genome Shotgun: Release 3 of the *Drosophila Melanogaster* Euchromatic Genome Sequence. *Genome Biology* **3** (12): RESEARCH0079. doi:10.1186/gb-2002-3-12-research0079.
- Chamary, J. V., J. L. Parmley, y L. D. Hurst. 2006. Hearing Silence: Non-Neutral Evolution at Synonymous Sites in Mammals. *Nature Reviews. Genetics* **7** (2): 98–108. doi:10.1038/nrg1770.
- Chan, A. H., P. A. Jenkins, y Y. S. Song. 2012. Genome-Wide Fine-Scale Recombination Rate Variation in *Drosophila Melanogaster*. *PLoS Genetics* **8** (12): e1003090. doi:10.1371/journal.pgen.1003090.
- Charlesworth, B., J. A. Coyne, y N. H. Barton. 1987. The Relative Rates of Evolution of Sex Chromosomes and Autosomes. *The American Naturalist* **130** (1): 113–146. doi:10.2307/2678832.
- Charlesworth, B., M. T. Morgan, y D. Charlesworth. 1993. The Effect of Deleterious Mutations on Neutral Molecular Variation. *Genetics* **134** (4): 1289–1303.
- Charlesworth, B. 1994. The Effect of Background Selection against Deleterious Mutations on Weakly Selected, Linked Variants. *Genetical Research* **63** (3): 213–27.
- Charlesworth, B., y D. Charlesworth. 2000. The Degeneration of Y Chromosomes. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* **355** (1403): 1563–1572. doi:10.1098/rstb.2000.0717.
- Charlesworth, B. 2001. The Effect of Life-History and Mode of Inheritance on Neutral Genetic Variability. *Genetical Research* **77** (2): 153–66.
- Charlesworth, J., y A. Eyre-Walker. 2006. The Rate of Adaptive Evolution in Enteric Bacteria. *Molecular Biology and Evolution* **23** (7): 1348–1356. doi:10.1093/molbev/msk025.
- Charlesworth, J., y A. Eyre-Walker. 2008. The McDonald-Kreitman Test and Slightly Deleterious Mutations. *Molecular Biology and Evolution* **25** (6): 1007–1015. doi:10.1093/molbev/msn005.
- Charlesworth, B. 2009. Fundamental Concepts in Genetics: Effective Population Size and Patterns of Molecular Evolution and Variation. *Nature Reviews. Genetics* **10** (3): 195–205. doi:10.1038/nrg2526.
- Charlesworth, B., y D. Charlesworth. 2009. Darwin and Genetics. *Genetics* **183** (3): 757–766. doi:10.1534/genetics.109.109991.
- Charlesworth, B. 2012. The Effects of Deleterious Mutations on Evolution at Linked Sites. *Genetics* **190** (1): 5–22. doi:10.1534/genetics.111.134288.
- Choi, K., X. Zhao, K. A. Kelly, O. Venn, J. D. Higgins, N. E. Yelina, T. J. Hardcastle, *et al.* 2014. Europe PMC Funders Group Arabidopsis Meiotic Crossover Hotspots Overlap with H2A.Z Nucleosomes at Gene Promoters. *Nat. Genet.* **45** (11): 1–24. doi:10.1038/ng.2766.Arabidopsis.
- Clemente, F., y C. Vogl. 2012. Evidence for Complex Selection on Four-Fold Degenerate Sites in *Drosophila Melanogaster*. *Journal of Evolutionary Biology* **25** (12): 2582–2595. doi:10.1111/jeb.12003.

- 
- Comeron, J. M., y M. Kreitman. 2000. The Correlation Between Intron Length and Recombination in *Drosophila*: Dynamic Equilibrium Between Mutational and Selective Forces. *Genetics* **156** (3): 1175–1190.
- Comeron, J. M. 2002. Population, Evolutionary and Genomic Consequences of Interference Selection. *Genetics* **161** (1): 389–410.
- Comeron, J. M., A. Williford, y R. M. Kliman. 2008. The Hill-Robertson Effect: Evolutionary Consequences of Weak Selection and Linkage in Finite Populations. *Heredity* **100** (1): 19–31. doi:10.1038/sj.hdy.6801059.
- Comeron, J. M., R. Ratnappan, y S. Bailin. 2012. The Many Landscapes of Recombination in *Drosophila Melanogaster*. *PLoS Genetics* **8** (10): e1002905. doi:10.1371/journal.pgen.1002905.
- Connallon, T. 2007. Adaptive Protein Evolution of X-Linked and Autosomal Genes in *Drosophila*: Implications for Faster-X Hypotheses. *Molecular Biology and Evolution* **24** (11): 2566–2572. doi:10.1093/molbev/msm199.
- Corbett-Detig, R. B., y D. L. Hartl. 2012. Population Genomics of Inversion Polymorphisms in *Drosophila Melanogaster*. *PLoS Genetics* **8** (12): e1003056. doi:10.1371/journal.pgen.1003056.
- Crawley, M. J. 2007. Regression, pág. 287–430, en *The R Book*. John Wiley & Sons, Chichester, West Sussex.
- Crick, F. H., L. Barnett, S. Brenner, y R. J. Watts-Tobin. 1961. General Nature of the Genetic Code for Proteins. *Nature* **192**: 1227–32.
- Cutter, A. D., y B. A. Payseur. 2003. Selection at Linked Sites in the Partial Selfer *Caenorhabditis Elegans*. *Molecular Biology and Evolution* **20** (5): 665–673. doi:10.1093/molbev/msg072.
- Cutter, A. D., y B. A. Payseur. 2013. Genomic Signatures of Selection at Linked Sites: Unifying the Disparity among Species. *Nature Reviews. Genetics* **14** (4): 262–74. doi:10.1038/nrg3425.
- Darwin, C. 1859. *On the Origin of Species by Means of Natural Selection. Or the Preservation of Favoured Races in the Struggle for Life*. John Murray, London.
- David, J., y P. Capy. 1988. Genetic Variation of *Drosophila Melanogaster* Natural Populations. *Trends in Genetics* **4** (4): 106–111. doi:10.1016/0168-9525(88)90098-4.
- Davison, A. C., y D. V. Hinkley. 1997. *Bootstrap Methods and Their Applications*. Cambridge University Press, Cambridge.
- Denver, D. R., P. C. Dolan, L. J. Wilhelm, W. Sung, J. I. Lucas-Lledó, D. K. Howe, S. C. Lewis, *et al.* 2009. A Genome-Wide View of *Caenorhabditis Elegans* Base-Substitution Mutation Processes. *Proc Natl Acad Sci USA* **106** (38): 16310–16314. doi:10.1073/pnas.0904895106.
- Denver, D. R., L. J. Wilhelm, D. K. Howe, K. Gafner, P. C. Dolan, y C. F. Baer. 2012. Variation in Base-Substitution Mutation in Experimental and Natural Lineages of *Caenorhabditis Nematodes*. *Genome Biology and Evolution* **4** (4): 513–522. doi:10.1093/gbe/evs028.
- Devlin, B., y N. Risch. 1995. A Comparison of Linkage Disequilibrium Measures for Fine-Scale Mapping. *Genomics* **29** (2): 311–322. doi:10.1159/000154430.
- Dobzhansky, T. G. 1937. *Genetics and the Origin of Species*. Columbia University Press, New York.
- Dobzhansky, T. G. 1955. A Review of Some Fundamental Concepts and Problems of Population Genetics. *Cold Spring Harbor Symposia on Quantitative Biology* **20**: 1–15.
- Drake, J. W. 1991. A Constant Rate of Spontaneous Mutation in DNA-Based Microbes. *Proc Natl Acad Sci USA* **88** (16): 7160–7164. doi:10.1073/pnas.88.16.7160.
- Duchen, P., D. Živković, S. Hutter, W. Stephan, y S. Laurent. 2013. Demographic Inference Reveals African and European Admixture in the North American *Drosophila Melanogaster* Population. *Genetics* **193** (1): 291–301. doi:10.1534/genetics.112.145912.

- 
- Elyashiv, E., K. Bullaughey, S. Sattath, Y. Rinott, M. Przeworski, y G. Sella. 2010. Shifts in the Intensity of Purifying Selection: An Analysis of Genome-Wide Polymorphism Data from Two Closely Related Yeast Species. *Genome Research* **20** (11): 1558–1573. doi:10.1101/gr.108993.110.
- Eöry, L., D. L. Halligan, y P. D. Keightley. 2010. Distributions of Selectively Constrained Sites and Deleterious Mutation Rates in the Hominid and Murid Genomes. *Molecular Biology and Evolution* **27** (1): 177–192. doi:10.1093/molbev/msp219.
- Eyre-Walker, A., y P. D. Keightley. 1999. High Genomic Deleterious Mutation Rates in Hominids. *Nature* **397** (6717): 344–347. doi:10.1038/16915.
- Eyre-Walker, A. 2006. The Genomic Rate of Adaptive Evolution. *Trends in Ecology & Evolution* **21** (10): 569–75. doi:10.1016/j.tree.2006.06.015.
- Eyre-Walker, A., M. Woolfit, y T. Phelps. 2006. The Distribution of Fitness Effects of New Deleterious Amino Acid Mutations in Humans. *Genetics* **173** (2): 891–900. doi:10.1534/genetics.106.057570.
- Eyre-Walker, A., y P. D. Keightley. 2007. The Distribution of Fitness Effects of New Mutations. *Nature Reviews. Genetics* **8** (8): 610–618. doi:10.1038/nrg2146.
- Eyre-Walker, A., y P. D. Keightley. 2009. Estimating the Rate of Adaptive Molecular Evolution in the Presence of Slightly Deleterious Mutations and Population Size Change. *Molecular Biology and Evolution* **26** (9): 2097–2108. doi:10.1093/molbev/msp119.
- Falconer, D. S., y T. F. C. Mackay. 1996. *Introduction to Quantitative Genetics*. 4th ed. Longman.
- Fay, J. C., y C. I. Wu. 2000. Hitchhiking under Positive Darwinian Selection. *Genetics* **155** (3): 1405–13.
- Fay, J. C., J. C. Fay, G. J. Wyckoff, G. J. Wyckoff, C. I. Wu, y C. I. Wu. 2001. Positive and Negative Selection on the Human Genome. *Genetics* **158** (3): 1227–34.
- Fay, J. C. 2011. Weighing the Evidence for Adaptation at the Molecular Level. *Trends in Genetics* **27** (9): 343–349. doi:10.1016/j.tig.2011.06.003.
- Fearnhead, P., y P. Donnelly. 2001. Estimating Recombination Rates from Population Genetic Data. *Genetics* **159** (3): 1299–1318.
- Feulner, P. G. D., F. J. J. Chain, M. Panchal, Y. Huang, C. Eizaguirre, M. Kalbe, T. L. Lenz, *et al.* 2015. Genomics of Divergence along a Continuum of Parapatric Population Differentiation. *PLOS Genetics* **11** (2): e1004966. doi:10.1371/journal.pgen.1004966.
- Filatov, D. A., y D. T. Gerrard. 2003. High Mutation Rates in Human and Ape Pseudoautosomal Genes. *Gene* **317** (1-2): 67–77. doi:10.1016/S0378-1119(03)00697-8.
- Fisher, R. A. 1930. *The Genetical Theory of Natural Selection*. The Clarendon Press, Oxford.
- Fiston-Lavier, A. S., N. D. Singh, M. Lipatov, y D. A. Petrov. 2010. Drosophila Melanogaster Recombination Rate Calculator. *Gene* **463** (1-2): 18–20. doi:10.1016/j.gene.2010.04.015.
- Foss, E., R. Lande, F. W. Stahl, y C. M. Steinberg. 1993. Chiasma Interference as a Function of Genetic-Distance (vol 133, Pg 681, 1993). *Genetics* **133** (3): 681–691.
- Fox, J. A., S. McMillan, y B. F. F. Ouellette. 2006. A Compilation of Molecular Biology Web Servers: 2006 Update on the Bioinformatics Links Directory. *Nucleic Acids Research* **34** (Web Server issue): 3–5. doi:10.1093/nar/gkl379.
- Fox, J. 2008. *Applied Regression Analysis and Generalized Linear Models*. 2nd ed, SAGE.
- Frisse, L., R. R. Hudson, A. Bartoszewicz, J. D. Wall, J. Donfack, y A. Di Rienzo. 2001. Gene Conversion and Different Population Histories May Explain the Contrast between Polymorphism and Linkage Disequilibrium Levels. *American Journal of Human Genetics* **69** (4): 831–843. doi:10.1086/323612.
- Galperin, M. Y. 2007. The Molecular Biology Database Collection: 2007 Update. *Nucleic Acids Research* **35** (Database issue: D3-4). doi:10.1093/nar/gkl1008.
- García-Dorado, A. 2012. Understanding and Predicting the Fitness Decline of Shrunken Populations: Inbreeding, Purging, Mutation, and Standard Selection. *Genetics* **190** (4): 1461–1476. doi:10.1534/genetics.111.135541.

- 
- Gay, J., S. Myers, y G. McVean. 2007. Estimating Meiotic Gene Conversion Rates from Population Genetic Data. *Genetics* **177** (2): 881–94. doi:10.1534/genetics.107.078907.
- Gillespie, J. H. 1991. *The Causes of Molecular Evolution*. Oxford University Press, New York.
- Gillespie, J. H. 2000. Genetic Drift in an Infinite Population: The Pseudohitchhiking Model. *Genetics* **155** (2): 909–919.
- Gillespie, J. H. 2001. Is the Population Size of a Species Relevant to Its Evolution? *Evolution; International Journal of Organic Evolution* **55** (11): 2161–2169. doi:10.1111/j.0014-3820.2001.tb00732.x.
- Glinka, S., L. Ometto, S. Mousset, W. Stephan, y D. De Lorenzo. 2003. Demography and Natural Selection Have Shaped Genetic Variation in *Drosophila Melanogaster*: A Multi-Locus Approach. *Genetics* **165** (3): 1269–1278.
- Goffeau, A., B. G. Barrell, H. Bussey, R. W. Davis, B. Dujon, H. Feldmann, F. Galibert, *et al.* 1995. Life with 6000 Genes. *Science* **274** (5287): 546,563-7.
- Gojobori, J., H. Tang, J. M. Akey, y C. Wu. 2007. Adaptive Evolution in Humans Revealed by the Negative Correlation between the Polymorphism and Fixation Phases of Evolution. *Proc Natl Acad Sci USA* **104** (10): 3907–3912. doi:10.1073/pnas.0605565104.
- Good, B. H., A. M. Walczak, R. A. Neher, y M. M. Desai. 2014. Genetic Diversity in the Interference Selection Limit. *PLoS Genetics* **10** (3): 1–14. doi:10.1371/journal.pgen.1004222.
- Gossmann, T. I., B. Song, A. J. Windsor, T. Mitchell-Olds, C. J. Dixon, M. V. Kapralov, D. A. Filatov, y A. Eyre-Walker. 2010. Genome Wide Analyses Reveal Little Evidence for Adaptive Evolution in Many Plant Species. *Molecular Biology and Evolution* **27** (8): 1822–1832. doi:10.1093/molbev/msq079.
- Gossmann, T. I., P. D. Keightley, y A. Eyre-Walker. 2012. The Effect of Variation in the Effective Population Size on the Rate of Adaptive Molecular Evolution in Eukaryotes. *Genome Biology and Evolution* **4** (5): 658–667. doi:10.1093/gbe/evs027.
- Gouy, M., y C. Gautier. 1982. Codon Usage in Bacteria: Correlation with Gene Expressivity. *Nucleic Acids Research* **10** (22): 7055–7074. doi:10.1093/nar/10.22.7055.
- Grantham, R., C. Gautier, M. Gouy, M. Jacobzone, y R. Mercier. 1981. Codon Catalog Usage Is a Genome Strategy Modulated for Gene Expressivity. *Nucleic Acids Research* **9** (1): 213. doi:10.1093/nar/9.1.213-b.
- Grossman, S. R., K. G. Andersen, I. Shlyakhter, S. Winnicki, A. Yen, D. J. Park, D. Griesemer, *et al.* 2014. Identifying Recent Adaptations in Large-Scale Genomic Data. *Cell* **152** (4): 703–713. doi:10.1016/j.cell.2013.01.035. Identifying.
- Hadany, L., y J. M. Comeron. 2008. Why Are Sex and Recombination so Common? *Annals of the New York Academy of Sciences* **1133**: 26–43. doi:10.1196/annals.1438.011.
- Hadrill, P. R., B. Charlesworth, D. L. Halligan, y P. Andolfatto. 2005. Patterns of Intron Sequence Evolution in *Drosophila* Are Dependent upon Length and GC Content. *Genome Biology* **6** (8): R67. doi:10.1186/gb-2005-6-8-r67.
- Hadrill, P. R., D. L. Halligan, D. Tomaras, y B. Charlesworth. 2007. Reduced Efficacy of Selection in Regions of the *Drosophila* Genome That Lack Crossing Over. *Genome Biology* **8** (2): R18. doi:10.1186/gb-2007-8-2-r18.
- Hadrill, P. R., K. R. Thornton, B. Charlesworth, y P. Andolfatto. 2005. Multilocus Patterns of Nucleotide Variability and the Demographic and Selection History of *Drosophila Melanogaster* Populations. *Genome Research* **15** (6): 790–799. doi:10.1101/gr.3541005.
- Hadrill, P. R., K. Zeng, y B. Charlesworth. 2011. Determinants of Synonymous and Nonsynonymous Variability in Three Species of *Drosophila*. *Molecular Biology and Evolution* **28** (5): 1731–1743. doi:10.1093/molbev/msq354.
- Hahn, M. W. 2007. Detecting Natural Selection on Cis-Regulatory DNA. *Genetica* **129** (1): 7–18. doi:10.1007/s10709-006-0029-y.
-

- 
- Hahn, M. W. 2008. Toward a Selection Theory of Molecular Evolution. *Evolution; International Journal of Organic Evolution* **62** (2): 255–65. doi:10.1111/j.1558-5646.2007.00308.x.
- Haldane, J.B.S. 1932. *The Causes of Evolution*. Longmans, Green and Co, London, New York, Toronto.
- Halligan, D. L., A. Eyre-Walker, P. Andolfatto, y P. D. Keightley. 2004. Patterns of Evolutionary Constraints in Intronic and Intergenic DNA of *Drosophila*. *Genome Research* **14** (2): 273–279. doi:10.1101/gr.1329204.
- Halligan, D. L., y P. D. Keightley. 2006. Ubiquitous Selective Constraints in the *Drosophila* Genome Revealed by a Genome-Wide Interspecies Comparison. *Genome Research* **16** (7): 875–884. doi:10.1101/gr.5022906.
- Halligan, D. L., F. Oliver, A. Eyre-Walker, B. Harr, y P. D. Keightley. 2010. Evidence for Pervasive Adaptive Protein Evolution in Wild Mice. *PLoS Genetics* **6** (1). doi:10.1371/journal.pgen.1000825.
- Harris, H. 1966. Enzyme Polymorphisms in Man. *Proceedings of the Royal Society of London. Series B, Containing Papers of a Biological Character. Royal Society (Great Britain)* **164** (1995): 298–310.
- Hartl, D. L., y A. G. Clark. 2007. *Principles of Population Genetics*. 4th ed. Sinauer Associates, Inc, Sunderland, Massachusetts.
- Hartl, D. L., E. N. Moriyama, y S. A. Sawyer. 1994. Selection Intensity for Codon Bias. *Genetics* **138** (1): 227–234. doi:10.3168/jds.S0022-0302(75)84789-8.
- Hellenthal, G., y M. Stephens. 2006. Insights into Recombination from Population Genetic Variation. *Current Opinion in Genetics & Development* **16** (6): 565–72. doi:10.1016/j.gde.2006.10.001.
- Hellmann, I., I. Ebersberger, S. E Ptak, S. Pääbo, y M. Przeworski. 2003. A Neutral Explanation for the Correlation of Diversity with Recombination Rates in Humans. *American Journal of Human Genetics* **72** (6): 1527–1535. doi:10.1086/375657.
- Hermisson, J., y P. S. Pennings. 2005. Soft Sweeps: Molecular Population Genetics of Adaptation from Standing Genetic Variation. *Genetics* **169** (4): 2335–52. doi:10.1534/genetics.104.036947.
- Hernandez, R. D., J. L. Kelley, E. Elyashiv, S. C. Melton, A. Auton, G. McVean, G. Sella, y M. Przeworski. 2011. Classic Selective Sweeps Were Rare in Recent Human Evolution. *Science* **331** (6019): 920–924. doi:10.1126/science.1198878.
- Hernandez, R. D. 2008. A Flexible Forward Simulator for Populations Subject to Selection and Demography. *Bioinformatics* **24** (23): 2786–2787. doi:10.1093/bioinformatics/btn522.
- Herr, D. G. 1986. On the History of ANOVA in Unbalanced, Factorial Designs: The First 30 Years. *The American Statistician* **40** (4): 265–270. doi:10.2307/2684597.
- Hershberg, R., y D. A. Petrov. 2009. General Rules for Optimal Codon Choice. *PLoS Genetics* **5** (7): e1000556. doi:10.1371/journal.pgen.1000556.
- Hey, J., y R. M. Kliman. 2002. Interactions between Natural Selection, Recombination and Gene Density in the Genes of *Drosophila*. *Genetics* **160** (2): 595–608.
- Hey, J., y J. Wakeley. 1997. A Coalescent Estimator of the Population Recombination Rate. *Genetics* **145** (3): 833–846.
- Hill, W. G., y B. S. Weir. 1994. Maximum-Likelihood Estimation of Gene Location by Linkage Disequilibrium. *American Journal of Human Genetics* **54** (4): 705–714.
- Hill, W. G., y A. Robertson. 1966. The Effect of Linkage on Limits to Artificial Selection. *Genetical Research* **8** (5–6): 269–294.
- Hill, W. G., y A. Robertson. 1968. The Effects of Inbreeding at Loci with Heterozygote Advantage. *Genetics* **60** (3): 615–628.
- Hilliker, A. J., S. H. Clark, y A. Chovnick. 1991. The Effect of DNA Sequence Polymorphisms on Intragenic Recombination in the Rosy Locus of *Drosophila Melanogaster*. *Genetics* **129** (3): 779–781.

- 
- Hilliker, A. J., G. Harauz, A. G. Reaume, M. Gray, S. H. Clark, A. Chovnick, y W. R. Engels. 1994. Meiotic Gene Conversion Tract Length Distribution within the Rosy Locus of *Drosophila Melanogaster*. *Genetics* **137** (4): 1019–1026.
- Hilliker, A.J., y A. Chovnick. 1981. Further Observations on Intragenic Recombination in *Drosophila Melanogaster*. *Genet Res.* **38** (2): 281–96.
- Hilton, H., R. M. Kliman, y J. Hey. 1994. Using Hitchhiking Genes to Study Adaptation and Divergence during Speciation within the *Drosophila-Melanogaster* Species Complex. *Evolution* **48** (6): 1900–1913. doi:10.2307/2410516.
- Hoskins, R A, J W Carlson, K H Wan, S Park, I Mendez, S E Galle, B W Booth, *et al.* 2015. The Release 6 Reference Sequence of the *Drosophila Melanogaster* Genome. *Genome Res.* **25** (3): 445–58. doi:10.1101/gr.185579.114.Freely.
- Hoskins, R. A., C. D. Smith, J. W. Carlson, A. B. Carvalho, A. Halpern, J. S. Kaminker, C. Kennedy, *et al.* 2002. Heterochromatic Sequences in a *Drosophila* Whole-Genome Shotgun Assembly. *Genome Biology* **3** (12). doi:10.1186/gb-2002-3-12-research0085.
- Hoskins, R. A., J. M. Landolin, J. B. Brown, J. E. Sandler, H. Takahashi, T. Lassmann, C. Yu, *et al.* 2011. Genome-Wide Analysis of Promoter Architecture in *Drosophila Melanogaster* **21** (2): 182–192. doi:10.1101/gr.112466.110.
- Howard, L. O. 1900. A Contribution to the Study of the Insect Fauna of Human Excrement. *Proc. Washington Acad. Sci.* **2**: 541–604.
- Huang, S. W., R. Friedman, N. Yu, A. Yu, y W. H. Li. 2005. How Strong Is the Mutagenicity of Recombination in Mammals? *Molecular Biology and Evolution* **22** (3): 426–431. doi:10.1093/molbev/msi025.
- Huang, W., A. Massouras, Y. Inoue, J. Peiffer, M. Ràmia, A. M. Tarone, L. Turlapati, *et al.* 2014. Natural Variation in Genome Architecture among 205 *Drosophila Melanogaster* Genetic Reference Panel Lines. *Genome Research* **24** (7): 1193–1208. doi:10.1101/gr.171546.113.
- Hudson, R. R. 1983. Properties of a Neutral Allele Model with Intragenic Recombination. *Theoretical Population Biology* **23** (2): 183–201. doi:10.1016/0040-5809(83)90013-8.
- Hudson, R. R., K. Bailey, D. Skarecky, J. Kwiatowski, y F. J. Ayala. 1994. Evidence for Positive Selection in the Superoxide Dismutase (Sod) Region of *Drosophila Melanogaster*. *Genetics* **136** (4): 1329–40.
- Hudson, R. R., M. Kreitman, y M. Aguadé. 1987. A Test of Neutral Molecular Evolution Based on Nucleotide Data. *Genetics* **116** (1): 153–9.
- Hudson, R. R. 2001. Two-Locus Sampling Distributions and Their Application. *Genetics* **159** (4): 1805–1817.
- Hughes, A. L., y M. Nei. 1988. Pattern of Nucleotide Substitution at Major Histocompatibility Complex Class I Loci Reveals Overdominant Selection. *Nature* **335** (6186): 167–70. doi:10.1038/335167a0.
- Hurst, L. D., y J. R. Peck. 1996. Recent Advances in Understanding of the Evolution and Maintenance of Sex. *Trends in Ecology & Evolution* **11** (2): 46–52. doi:10.1016/0169-5347(96)81041-X.
- Hutter, S., A. J. Vilella, y J. Rozas. 2006. Genome-Wide DNA Polymorphism Analyses Using VariScan. *BMC Bioinformatics* **7**: 409. doi:10.1186/1471-2105-7-409.
- Huynh, L. Y., D. L. Maney, y J. W. Thomas. 2011. Chromosome-Wide Linkage Disequilibrium Caused by an Inversion Polymorphism in the White-Throated Sparrow (*Zonotrichia Albicollis*). *Heredity* **106** (4): 537–546. doi:10.1038/hdy.2010.85.
- Ikemura, T. 1981. Correlation between the Abundance of *Escherichia Coli* Transfer RNAs and the Occurrence of the Respective Codons in Its Protein Genes: A Proposal for a Synonymous Codon Choice That Is Optimal for the *E. Coli* Translational System. *Journal of Molecular Biology* **151** (3): 389–409. doi:10.1016/0022-2836(81)90363-6.
-

- 
- Ivanov, A. I., A. C. Rovescalli, P. Pozzi, S. Yoo, B. Mozer, H. Li, S. Yu, *et al.* 2004. Genes Required for Drosophila Nervous System Development Identified by RNA Interference. *Proc Natl Acad Sci USA* **101** (46): 16216–21. doi:10.1073/pnas.0407188101.
- Jeffreys, A. J., y C. A. May. 2004. Intense and Highly Localized Gene Conversion Activity in Human Meiotic Crossover Hot Spots. *Nature Genetics* **36** (2): 151–6. doi:10.1038/ng1287.
- Jensen-seaman, M. I., T. S. Furey, B. A. Payseur, Y. Lu, K. M. Roskin, C. Chen, M. A. Thomas, D. Haussler, y H. J. Jacob. 2004. Comparative Recombination Rates in the Rat, Mouse, and Human Genomes **14** (4): 528–538. doi:10.1101/gr.1970304.1.
- Johnson, C. W. 1913. The Distribution of Some Species of Drosophila. *Psyche: A Journal of Entomology* **20** (6): 202–205. doi:10.1155/1913/41505.
- Johnson, F. M., C. G. Kanapi, R. H. Richardson, M. R. Wheeler, y W. S. Stone. 1966. An Analysis of Polymorphisms among Isozyme Loci in Dark and Light Drosophila Ananassae Strains from American and Western Samoa. *Proc Natl Acad Sci USA* **56** (1): 119–25.
- Jorde, L. B., W. S. Watkins, y M. J. Bamshad. 2001. Population Genomics: A Bridge from Evolutionary History to Genetic Medicine. *Human Molecular Genetics* **10** (20): 2199–2207. doi:10.1093/hmg/10.20.2199.
- Jukes, T. H., y C. R. Cantor. 1969. Mammalian Protein Metabolism, pág. 21-132 en *Evolution of Protein Molecules*, editado por HN Munro. Academic Press, New York.
- Kaminker, J., C. Bergman, B. Kronmiller, J. Carlson, R. Svirskas, S. Patel, E. Frise, *et al.* 2002. The Transposable Elements of the Drosophila Melanogaster Euchromatin: A Genomics Perspective. *Genome Biology* **3** (12): RESEARCH0084. doi:10.1186/gb-2002-3-12-research0084.
- Kaplan, N. L., R. R. Hudson, y C. H. Langley. 1989. The ‘Hitchhiking Effect’ Revisited. *Genetics* **123** (4): 887–99.
- Keeney, S. 2001. Mechanism and Control of Meiotic Recombination Initiation. *Current Topics in Developmental Biology* **52**: 1-53.
- Keightley, P. D., y A. Eyre-Walker. 2007. Joint Inference of the Distribution of Fitness Effects of Deleterious Mutations and Population Demography Based on Nucleotide Polymorphism Frequencies. *Genetics* **177** (4): 2251–2261. doi:10.1534/genetics.107.080663.
- Keightley, P. D., U. Trivedi, M. Thomson, F. Oliver, S. Kumar, y M. L. Blaxter. 2009. Analysis of the Genome Sequences of Three Drosophila Melanogaster Spontaneous Mutation Accumulation Lines. *Genome Research* **19** (7): 1195–1201. doi:10.1101/gr.091231.109.
- Keightley, P. D., y A. Eyre-Walker. 2010. What Can We Learn about the Distribution of Fitness Effects of New Mutations from DNA Sequence Data? *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* **365** (1544): 1187–1193. doi:10.1098/rstb.2009.0266.
- Keightley, P. D. 2012. Estimating the Rate of Adaptive Molecular Evolution When the Evolutionary Divergence between Species Is Small. *Journal of Molecular Evolution* **74** (1-2): 61–68. doi:10.1007/s00239-012-9488-1.
- Keller, A. 2007. Drosophila Melanogaster’s History as a Human Commensal. *Current Biology* **17** (3): R77–R81. doi:10.1016/j.cub.2006.12.031.
- Kim, Y., y R. Neilsen. 2004. Linkage Disequilibrium as a Signature of Selective Sweeps. *Genetics* **167** (3): 1513–1524. doi:10.1534/genetics.103.025387.
- Kim, Y., y W. Stephan. 2002. Detecting a Local Signature of Genetic Hitchhiking along a Recombining Chromosome. *Genetics* **160** (2): 765–77.
- Kimura, M., y T. Ohta. 1969. The Average Number of Generations until Fixation of a Mutant Gene in a Finite Population. *Genetics* **61** (3): 763–71.
- Kimura, M. 1968. Evolutionary Rate at the Molecular Level. *Nature* **217** (5129): 624–626. doi:10.1038/217624a0.



- 
- Kimura, M. 1983. *The Neutral Theory of Molecular Evolution*. Cambridge University Press, Cambridge, Massachusetts.
- King, J. L., y T. H. Jukes. 1969. Non-Darwinian Evolution. *Science* **164** (3881): 788–798. doi:10.1126/science.164.3881.788.
- Kingman, J. F. C. 1982a. On the Genealogy of Large Populations. *Journal of Applied Probability* **19**: 27–43.
- Kingman, J. F. C. 1982b. The Coalescent. *Stochastic Processes and Their Applications* **13** (3): 235–248. doi:10.1016/0304-4149(82)90011-4.
- Kong, A., D. F. Gudbjartsson, J. Sainz, G. M. Jonsdottir, S. A. Gudjonsson, B. Richardsson, S. Sigurdardottir, et al. 2002. A High-Resolution Recombination Map of the Human Genome. *Nature Genetics* **31** (3): 241–247. doi:10.1038/ng917.
- Kopczynski, C. C., J. N. Noordermeer, T. L. Serano, W. Y. Chen, J. D. Pendleton, S. Lewis, C. S. Goodman, y G. M. Rubin. 1998. A High Throughput Screen to Identify Secreted and Transmembrane Proteins Involved in *Drosophila* Embryogenesis. *Proc Natl Acad Sci USA* **95** (17): 9973–9978. doi:10.1073/pnas.95.17.9973.
- Kreitman, M. 1983. Nucleotide Polymorphism at the Alcohol Dehydrogenase Locus of *Drosophila Melanogaster*. *Nature* **304** (5925): 412–417.
- Kreitman, M. 2000. Methods to Detect Selection in Populations with Applications to the Human. *Annual Review of Genomics and Human Genetics* **1**: 539–59. doi:10.1146/annurev.genom.1.1.539.
- Kulathinal, R. J., L. S. Stevison, y M. A. F. Noor. 2009. The Genomics of Speciation in *Drosophila*: Diversity, Divergence, and Introgression Estimated Using Low-Coverage Genome Sequencing. *PLoS Genetics* **5** (7). doi:10.1371/journal.pgen.1000550.
- Künstner, A., B. Nabholz, y H. Ellegren. 2011. Significant Selective Constraint at 4-Fold Degenerate Sites in the Avian Genome and Its Consequence for Detection of Positive Selection. *Genome Biology and Evolution* **3** (1): 1381–1389. doi:10.1093/gbe/evr112.
- Laayouni, H., L. Montanucci, M. Sikora, M. Melé, G. M. Dall’Olio, B. Lorente-Galdos, K. M. McGee, et al. 2011. Similarity in Recombination Rate Estimates Highly Correlates with Genetic Differentiation in Humans. *PLoS ONE* **6** (3). doi:10.1371/journal.pone.0017913.
- Lachaise, D., M. Cairou, J. R. David, F. Lemeunier, L. Tsacas, y M. Ashburner. 1988. Historical Biogeography of *Drosophila Melanogaster* Species Subgroup. *Evolutionary Biology* **22**: 159–225.
- Langley, C. H., B. P. Lazzaro, W. Phillips, E. Heikkinen, y J. M. Braverman. 2000. Linkage Disequilibria and the Site Frequency Spectra in the Su(s) and Su(wa) Regions of the *Drosophila Melanogaster* X Chromosome. *Genetics* **156** (4): 1837–1852.
- Langley, C. H, K. Stevens, C. Cardeno, Y. C. G. Lee, D. R. Schrider, J. E. Pool, S. A. Langley, et al. 2012. Genomic Variation in Natural Populations of *Drosophila Melanogaster*. *Genetics* **192** (2): 533–98. doi:10.1534/genetics.112.142018.
- Langsrud, Øyvind. 2003. ANOVA for Unbalanced Data: Use Type II instead of Type III Sums of Squares. *Statistics and Computing* **13** (1969): 163–167. doi:10.1023/A:1023260610025.
- Larracuente, A. M., T. B. Sackton, A. J. Greenberg, A. Wong, N. D. Singh, D. Sturgill, Y. Zhang, B. Oliver, y A. G. Clark. 2008. Evolution of Protein-Coding Genes in *Drosophila*. *Trends in Genetics* **24** (3): 114–123. doi:10.1016/j.tig.2007.12.001.
- Lawrie, D. S., P. W. Messer, R. Hershberg, y D. A. Petrov. 2013. Strong Purifying Selection at Synonymous Sites in *D. Melanogaster*. *PLoS Genetics* **9** (5): 33–40. doi:10.1371/journal.pgen.1003527.
- Lee, H., E. Popodi, H. Tang, y P. L. Foster. 2012. PNAS Plus: Rate and Molecular Spectrum of Spontaneous Mutations in the Bacterium *Escherichia Coli* as Determined by Whole-Genome Sequencing. *Proceedings of the National Academy of Sciences* **109** (41): E2774–E2783. doi:10.1073/pnas.1210309109.
- Lercher, M. J., y L. D. Hurst. 2002. Human SNP Variability and Mutation Rate Are Higher in Regions of High Recombination. *Trends in Genetics* **18** (7): 337–340. doi:10.1016/S0168-9525(02)02669-0.
-

- 
- Lewin, B. 1994. *Genes V*. Oxford University Press.
- Lewis, E. B. 1978. A Gene Complex Controlling Segmentation in *Drosophila*. *Nature* **276** (5688): 565–570. doi:10.1038/276565a0.
- Lewontin, R. C. 1964. The Interaction of Selection and Linkage. I. General Considerations; Heterotic Models. *Genetics* **49** (1): 49–67.
- Lewontin, R. C., y J. L. Hubby. 1966. A Molecular Approach to the Study of Genic Heterozygosity in Natural Populations. II. Amount of Variation and Degree of Heterozygosity in Natural Populations of *Drosophila Pseudoobscura*. *Genetics* **54** (2): 595–609.
- Lewontin, R. C., y J. Krakauer. 1973. Distribution of Gene Frequency as a Test of the Theory of the Selective Neutrality of Polymorphisms. *Genetics* **74** (1): 175–95.
- Lewontin, R. C. 1974. *The Genetic Basis of Evolutionary Change*. Columbia University Press, New York.
- Li, H., y W. Stephan. 2006. Inferring the Demographic History and Rate of Adaptive Substitution in *Drosophila*. *PLoS Genetics* **2** (10): 1580–1589. doi:10.1371/journal.pgen.0020166.
- Li, W. H. 1993. Unbiased Estimation of the Rates of Synonymous and Nonsynonymous Substitution. *Journal of Molecular Evolution* **36** (1): 96–99. doi:10.1007/BF02407308.
- Li, W. H., Z. Gu, H. Wang, y A. Nekrutenko. 2001. Evolutionary Analyses of the Human Genome. *Nature* **409**: 847–9. doi:10.1038/35057039.
- Li, W. H., C. I. Wu, y C. C. Luo. 1985. A New Method for Estimating Synonymous and Nonsynonymous Rates of Nucleotide Substitution Considering the Relative Likelihood of Nucleotide and Codon Changes. *Mol. Biol. Evol.* **2** (2): 150–174.
- Lichten, M. 2001. Meitoci Recombination: Breaking the Genome to Save It. *Current Biology* **11**: 253–256.
- Liti, G., D. M. Carter, A. M. Moses, J. Warringer, L. Parts, S. A. James, R. P. Davey, *et al.* 2009. Population Genomics of Domestic and Wild Yeasts **458** (7236): 337–341. doi:10.1038/nature07743.Population.
- Litt, M., y L. B. Jorde. 1986. Linkage Disequilibria between Pairs of Loci within a Highly Polymorphic Region of Chromosome 2Q. *American Journal of Human Genetics* **39** (2): 166–178.
- Lohmueller, K. E., A. Albrechtsen, Y. Li, S. Y. Kim, T. Korneliussen, N. Vinckenbosch, G. Tian, *et al.* 2011. Natural Selection Affects Multiple Aspects of Genetic Variation at Putatively Neutral Sites across the Human Genome. *PLoS Genetics* **7** (10): e1002326. doi:10.1371/journal.pgen.1002326.
- Lynch, Michael. 2006. The Origins of Eukaryotic Gene Structure. *Molecular Biology and Evolution* **23** (2): 450–468. doi:10.1093/molbev/msj050.
- Lynch, Michael. 2007. *The Origins of Genome Architecture*. Sinauer Associates, Inc., Sunderland, Massachusetts.
- Lynch, M., W. Sung, K. Morris, N. Coffey, C. R. Landry, E. B. Dopman, W. J. Dickinson, *et al.* 2008. A Genome-Wide View of the Spectrum of Spontaneous Mutations in Yeast. *Proc Natl Acad Sci USA* **105** (27): 9272–9277. doi:10.1073/pnas.0803466105.
- Machado, C. A., R. M. Kliman, J. A. Markert, y J. Hey. 2002. Inferring the History of Speciation from Multilocus DNA Sequence Data: The Case of *Drosophila Pseudoobscura* and Close Relatives. *Molecular Biology and Evolution* **19** (4): 472–488. doi:10.1093/oxfordjournals.molbev.a004103.
- Mackay, T. F. C., S. Richards, E. A. Stone, A. Barbadilla, J. F. Ayroles, D. Zhu, S. Casillas, *et al.* 2012. The *Drosophila Melanogaster* Genetic Reference Panel. *Nature* **482** (7384): 173–8. doi:10.1038/nature10811.
- Manku, H., C. D. Langefeld, S. G. Guerra, T. H. Malik, M. Alarcon-Riquelme, J. M. Anaya, S. C. Bae, *et al.* 2013. Trans-Ancestral Studies Fine Map the SLE-Susceptibility Locus TNFSF4. *PLoS Genetics* **9** (7): e1003554. doi:10.1371/journal.pgen.1003554.

- 
- Manolio, T. A., F. S. Collins, N. J. Cox, D. B. Goldstein, L. A. Hindorff, D. J. Hunter, M. I. McCarthy, *et al.* 2009. Finding the Missing Heritability of Complex Diseases. *Nature* **461** (7265): 747–753. doi:10.1038/nature08494.
- Marais, G., D. Mouchiroud, y L. Duret. 2001. Does Recombination Improve Selection on Codon Usage? Lessons from Nematode and Fly Complete Genomes. *Proc Natl Acad Sci USA* **98** (10): 5688–5692. doi:10.1073/pnas.091427698.
- Matthews, K. A., T. C. Kaufman, y W. M. Gelbart. 2005. Research Resources for Drosophila: The Expanding Universe. *Nature Reviews. Genetics* **6** (3): 179–193. doi:10.1038/nrg1554.
- Mayr, E. 1942. *Systematics and the Origin of Species, from the Viewpoint of a Zoologist*. Columbia University Press, New York.
- Mayr, E. 1963. *Animal Species and Evolution*. Cambridge, Massachusetts: Harvard University Press.
- Mayr, E. 1976. *Evolution and Diversity of Life: Selected Essays*. Harvard University Press, Cambridge, Massachusetts.
- Mayr, E. 1991. *One Long Argument: Charles Darwin and the Genesis of Modern Evolutionary Thought*. Cambridge, Massachusetts: Harvard University Press.
- McDonald, J. H., y M. Kreitman. 1991. Adaptive Protein Evolution at the Adh Locus in Drosophila. *Nature* **351** (6328): 652–4. doi:10.1038/351652a0.
- McGaugh, S. E., C. S. S. Heil, B. Manzano-Winkler, L. Loewe, S. Goldstein, T. L. Himmel, y M. A. F. Noor. 2012. Recombination Modulates How Selection Affects Linked Sites in Drosophila. *PLoS Biology* **10** (11): e1001422. doi:10.1371/journal.pbio.1001422.
- McVean, G. 2007. The Structure of Linkage Disequilibrium around a Selective Sweep. *Genetics* **175** (3): 1395–1406. doi:10.1534/genetics.106.062828.
- McVean, G., P. Awadalla, y P. Fearnhead. 2002. A Coalescent-Based Method for Detecting and Estimating Recombination from Gene Sequences. *Genetics* **160** (3): 1231–1241.
- McVean, G. A. T., y B. Charlesworth. 2000. The Effects of Hill-Robertson Interference between Weakly Selected Mutations on Patterns of Molecular Evolution and Variation. *Genetics* **155** (2): 929–944.
- Mcvean, G. A. T., S. R. Myers, S. Hunt, P. Deloukas, D. R. Bentley, y P. Donnelly. 2004. The Fine-Scale Structure of Recombination Rate Variation in the Human Genome. *Science* **304** (5670): 581–4. doi:10.1126/science.
- McVean, G. A. T., y J. Vieira. 2001. Inferring Parameters of Mutation, Selection and Demography from Patterns of Synonymous Site Evolution in Drosophila. *Genetics* **157** (1931): 245–257.
- Messer, P. W., y D. A. Petrov. 2013. Frequent Adaptation and the McDonald-Kreitman Test. *Proc Natl Acad Sci USA* **110** (21): 8615–20. doi:10.1073/pnas.1220835110.
- Mikkelsen, T. S., L. W. Hillier, E. E. Eichler, M. C. Zody, D. B. Jaffe, *et al.* 2005. Initial Sequence of the Chimpanzee Genome and Comparison with the Human Genome. *Nature* **437** (7055): 69–87. doi:10.1038/nature04072.
- Miller, D. E., S. Takeo, K. Nandanan, A. Paulson, M. M. Gogol, A. C. Noll, A. G. Perera, *et al.* 2012. A Whole-Chromosome Analysis of Meiotic Recombination in Drosophila Melanogaster. *G3; Genes/Genomes/Genetics* **2** (2): 249–260. doi:10.1534/g3.111.001396.
- Misra, S., M. A. Crosby, C. J. Mungall, B. B. Matthews, K. S. Campbell, P. Hradecky, Y. Huang, *et al.* 2002. Annotation of the Drosophila Melanogaster Euchromatic Genome: A Systematic Review. *Genome Biology* **3** (12): RESEARCH0083. doi:10.1186/gb-2002-3-12-research0083.
- Morgan, T. H. 1911. Random Segregation versus Coupling in Mendelian Inheritance. *Science* **33** (873): 384.
- Morgan, T. H. 1912. Complete Linkage in the Second Chromosome of the Male of Dorosophila. *Science* **36**: 719–720.
-

- 
- Morgan, T. H., Sturtevant AH, Muller HJ, Bridges CB. 1915. The Mechanism of Mendelian Heredity. Henry Holt & Company, New York.
- Morris, G. P., P. Ramu, S. P. Deshpande, C. T. Hash, T. Shah, H. D. Upadhyaya, O. Riera-Lizarazu, *et al.* 2013. Population Genomic and Genome-Wide Association Studies of Agroclimatic Traits in Sorghum. *Proc Natl Acad Sci USA* **110** (2): 453–8. doi:10.1073/pnas.1215985110.
- Mortimer, R. K., y S. Fogel. 1974. Genetical Interference and Gene Conversion, pág. 263–275 en *Mechanism in Recombination*, editado por RF Grell, New York.
- Muggeo, V. M. R. 2003. Estimating Regression Models with Unknown Break-Points. *Statistics in Medicine* **22** (19): 3055–3071. doi:10.1002/sim.1545.
- Muller, H. J., y W. D. Kaplan. 1966. The Dosage Compensation of Drosophila and Mammals as Showing the Accuracy of the Normal Type. *Genetical Research* **8** (01): 41–59. doi:10.1017/S0016672300009903.
- Muller, H. J. 1927. Artificial Transmutation of the Gene. *Science* **66** (1699): 84–87.
- Nachman, M. W. 2001. Single Nucleotide Polymorphisms and Recombination Rate in Humans. *Trends in Genetics : TIG* **17** (9): 481–485. doi:10.1016/S0168-9525(01)02409-X.
- Nachman, M. W. 1997. Patterns of DNA Variability at X-Linked Loci in Mus Domesticus. *Genetics* **147** (3): 1303–1316.
- Navarro, A., E. Betrán, A. Barbadilla, y A. Ruiz. 1996. Dynamics of Gametic Disequilibria between Loci Linked to Chromosome Inversions: The Recombination Redistributing Effect of Inversions. *Genetical Research* **67** (1): 67–76.
- Navarro, A., A. Barbadilla, y A. Ruiz. 2000. Effect of Inversion Polymorphism on the Neutral Nucleotide Variability of Linked Chromosomal Regions in Drosophila. *Genetics* **155** (2): 685–698.
- Nègre, N., C. D. Brown, L. Ma, C. A. Bristow, W. Miller, U. Wagner, P. Kheradpour, *et al.* 2011. A Cis-Regulatory Map of the Drosophila Genome. *Nature* **471** (7339): 527–531. doi:10.1038/nature09990.A.
- Neher, R. A., T. A. Kessinger, y B. I. Shraiman. 2013. Coalescence and Genetic Diversity in Sexual Populations under Selection. *Proc Natl Acad Sci USA* **110** (39): 15836–41. doi:10.1073/pnas.1309697110..
- Nei, M. 1987. *Molecular Evolutionary Genetics*, Columbia University Press, New York.
- Nevo, E., A. Beiles, y R. Ben-Shlomo. 1984. The Evolutionary Significance of Genetic Diversity: Ecological, Demographic and Life History Correlates, pág. 13–213 en *Evolutionary Dynamics of Genetic Diversity. Proceedings of a Symposium Held in Manchester, England, March 29–30, 1983*, editado por G. S. Mani. Springer-Verlag, Berlin Heidelberg. doi:10.1007/978-3-642-51588-0\_2.
- Nielsen, R. 2001. Statistical Tests of Selective Neutrality in the Age of Genomics. *Heredity* **86** (Pt 6): 641–7.
- Nielsen, R., y Z. Yang. 1998. Likelihood Models for Detecting Positively Selected Amino Acid Sites and Applications to the HIV-1 Envelope Gene. *Genetics* **148** (3): 929–36.
- Nielsen, R., V. L. Bauer DuMont, M. J. Hubisz, y C. F. Aquadro. 2007. Maximum Likelihood Estimation of Ancestral Codon Usage Bias Parameters in Drosophila. *Molecular Biology and Evolution* **24** (1): 228–235. doi:10.1093/molbev/msl146.
- Nielsen, R., C. Bustamante, A. G. Clark, S. Glanowski, T. B. Sackton, M. J. Hubisz, A. Fledel-Alon, *et al.* 2005. A Scan for Positively Selected Genes in the Genomes of Humans and Chimpanzees. *PLoS Biology* **3** (6): e170. doi:10.1371/journal.pbio.0030170.
- Nielsen, R., y Z. Yang. 2003. Estimating the Distribution of Selection Coefficients from Phylogenetic Data with Applications to Mitochondrial and Viral DNA. *Molecular Biology and Evolution* **20** (8): 1231–1239. doi:10.1093/molbev/msg147.
- Nüsslein-Volhard, C., y E. Wieschaus. 1980. Mutations Affecting Segment Number and Polarity in Drosophila. *Nature* **287** (5785): 795–801. doi:10.1038/287795a0.

- 
- Ohta, T. 2002. Near-Neutrality in Evolution of Genes and Gene Regulation. *Proc Natl Acad Sci USA* **99** (25): 16134–16137. doi:10.1073/pnas.252626899.
- Olson-Manning, C. F., M. R. Wagner, y T. Mitchell-Olds. 2012. Adaptive Evolution: Evaluating Empirical Support for Theoretical Predictions. *Nature Reviews. Genetics* **13** (12): 867–77. doi:10.1038/nrg3322.
- Orr, H. A. 2005. The Genetic Theory of Adaptation: A Brief History. *Nature Reviews. Genetics* **6** (2): 119–127. doi:10.1038/nrg1523.
- Orr, H. A., y A. J. Betancourt. 2001. Haldane’s Sieve and Adaptation from the Standing Genetic Variation. *Genetics* **157** (1977): 875–884.
- Ossowski, S., K. Schneeberger, J. I. Lucas-Lledó, N. Warthmann, R. M. Clark, R. G. Shaw, D. Weigel, y M. Lynch. 2010. The Rate and Molecular Spectrum of Spontaneous Mutations in Arabidopsis Thaliana. *Science (New York, N.Y.)* **327** (5961): 92–94. doi:10.1126/science.1180677.
- Paape, T., P. Zhou, A. Branca, R. Briskine, N. Young, y P. Tiffin. 2012. Fine-Scale Population Recombination Rates, Hotspots, and Correlates of Recombination in the Medicago Truncatula Genome. *Genome Biology and Evolution* **4** (5): 726–737. doi:10.1093/gbe/evs046.
- Pang, A. W., J. R. MacDonald, D. Pinto, J. Wei, M. A. Rafiq, D. F. Conrad, H. Park, *et al.* 2010. Towards a Comprehensive Structural Variation Map of an Individual Human Genome. *Genome Biology* **11** (5): R52. doi:10.1186/gb-2010-11-5-r52.
- Parsch, J., S. Novozhilov, S. S. Saminadin-Peter, K. M. Wong, y P. Andolfatto. 2010. On the Utility of Short Intron Sequences as a Reference for the Detection of Positive and Negative Selection in Drosophila. *Molecular Biology and Evolution* **27** (6): 1226–1234. doi:10.1093/molbev/msq046.
- Pennings, P. S., y J. Hermisson. 2006. Soft Sweeps II--Molecular Population Genetics of Adaptation from Recurrent Mutation or Migration. *Molecular Biology and Evolution* **23** (5): 1076–84. doi:10.1093/molbev/msj117.
- Pink, C. J., S. K. Swaminathan, I. Dunham, J. Rogers, A. Ward, y L. D. Hurst. 2009. Evidence That Replication-Associated Mutation Alone Does Not Explain between-Chromosome Differences in Substitution Rates. *Genome Biology and Evolution* **1**: 13–22. doi:10.1093/gbe/evp001.
- Plotkin, J. B., y G. Kudla. 2011. Synonymous but Not the Same: The Causes and Consequences of Codon Bias. *Nature Reviews. Genetics* **12** (1): 32–42. doi:10.1038/nrg2899.
- Pollard, K. S., M. J. Hubisz, K. R. Rosenbloom, y A. Siepel. 2010. Detection of Nonneutral Substitution Rates on Mammalian Phylogenies. *Genome Research* **20** (1): 110–121. doi:10.1101/gr.097857.109.
- Pool, J. E., R. B. Corbett-Detig, R. P. Sugino, K. A. Stevens, C. M. Cardeno, M. W. Crepeau, P. Duchon, *et al.* 2012. Population Genomics of Sub-Saharan Drosophila Melanogaster: African Diversity and Non-African Admixture. *PLoS Genetics* **8** (12): e1003080. doi:10.1371/journal.pgen.1003080.
- Pool, J. E., y R. Nielsen. 2007. Population Size Changes Reshape Genomic Patterns of Diversity. *Evolution* **61** (12): 3001–3006. doi:doi:10.1111/j.1558-5646.2007.00238.x.
- Pool, J. E., y R. Nielsen. 2008. The Impact of Founder Events on Chromosomal Variability in Multiply Mating Species. *Molecular Biology and Evolution* **25** (8): 1728–1736. doi:10.1093/molbev/msn124.
- Powell, J. R. 1997. *Progress and Prospects in Evolutionary Biology: The Drosophila Model*. Oxford University Press, New York.
- Presgraves, D. C. 2005. Recombination Enhances Protein Adaptation in Drosophila Melanogaster. *Current Biology* **15** (18): 1651–1656. doi:10.1016/j.cub.2005.07.065.
- Ptak, S. E., K. Voelpel, y M. Przeworski. 2004. Insights into Recombination from Patterns of Linkage Disequilibrium in Humans. *Genetics* **167** (1): 387–397. doi:10.1534/genetics.167.1.387.
- Ràmia, M., P. Librado, S. Casillas, J. Rozas, y A. Barbadilla. 2012. PopDrowser: The Population Drosophila Browser. *Bioinformatics* **28** (4): 595–596. doi:10.1093/bioinformatics/btr691.
-

- 
- Reed, F. A., y S. A. Tishkoff. 2006. Positive Selection Can Create False Hotspots of Recombination. *Genetics* **172** (3): 2011–2014. doi:10.1534/genetics.105.052183.
- Roberts, D. B. 2006. *Drosophila Melanogaster*: The Model Organism. *Entomologia Experimentalis et Applicata* **121** (2): 93–103. doi:10.1111/j.1570-8703.2006.00474.x.
- Romiguier, J., P. Gayral, M. Ballenghien, A. Bernard, V. Cahais, A. Chenuil, Y. Chiari, *et al.* 2014. Comparative Population Genomics in Animals Uncovers the Determinants of Genetic Diversity. *Nature* **515** (7526):261-3. doi:10.1038/nature13685.
- Rubin, G. M. 1966. Around the Genomes: The Drosophila Genoma Project. *Genome Research* **6** (2): 71–79.
- Russo, C. A. N. Takezaki, y M. Nei. 1995. Molecular Phylogeny and Divergence Times of Drosophilid Species. *Molecular Biology and Evolution* **12** (3): 391–404.
- Sabeti, P. C., S. F. Schaffner, B. Fry, J. Lohmueller, P. Varilly, O. Shamovsky, A. Palma, T. S. Mikkelsen, D. Altshuler, y E. S. Lander. 2006. Positive Natural Selection in the Human Lineage. *Science* **312** (5780): 1614–20. doi:10.1126/science.1124309.
- Sabeti, P. C., D. E. Reich, J. M. Higgins, H. Z. P. Levine, D. J. Richter, S. F. Schaffner, S. B. Gabriel, *et al.* 2002. Detecting Recent Positive Selection in the Human Genome from Haplotype Structure. *Nature* **419** (6909): 832–7. doi:10.1038/nature01140.
- Sabeti, P. C., P. Varilly, B. Fry, J. Lohmueller, E. Hostetter, C. Cotsapas, X. Xie, *et al.* 2009. Genome-Wide Detection and Characterization of Positive Selection in Human Populations. *Nature* **449** (7164): 913–918. doi:10.1038/nature06250.
- Sackton, T. B., R. J. Kulathinal, C. M. Bergman, A. R. Quinlan, E. B. Dopman, M. Carneiro, G. T. Marth, D. L. Hartl, y A. G. Clark. 2009. Population Genomic Inferences from Sparse High-Throughput Sequencing of Two Populations of *Drosophila Melanogaster*. *Genome Biology and Evolution* **1**: 449–465. doi:10.1093/gbe/evp048.
- Sawyer, S. A., y D. L. Hartl. 1992. Population Genetics of polymorphism and divergence. *Genetics* **132** (4): 1161–1176.
- Schneider, A., B. Charlesworth, A. Eyre-Walker, y P. D. Keightley. 2011. A Method for Inferring the Rate of Occurrence and Fitness Effects of Advantageous Mutations. *Genetics* **189** (4): 1427–1437. doi:10.1534/genetics.111.131730.
- Schrider, D. R., D. Houle, M. Lynch, y M. W. Hahn. 2013. Rates and Genomic Consequences of Spontaneous Mutational Events in *Drosophila Melanogaster*. *Genetics* **194** (4): 937–954. doi:10.1534/genetics.113.151670.
- Shapiro, J. A., W. Huang, C. Zhang, M. J. Hubisz, J. Lu, D. A. Turissini, S. Fang, *et al.* 2007. Adaptive Genic Evolution in the *Drosophila* Genomes. *Proc Natl Acad Sci USA* **104** (7): 2271–6. doi:10.1073/pnas.0610385104.
- Sharp, P. M., K. M. Devine. 1989. Codon Usage and Gene Expression Level in *Dictyosteiium Discooidtum*: Highly Expressed Genes Do Prefer Optimal Codons. *Nucleic Acids Res* **17** (13): 5029–5040.
- Sharp, P. M., y W. H. Li. 1989. On the Rate of DNA Sequence Evolution in *Drosophila*. *Journal of Molecular Evolution* **28** (5): 398–402.
- Simpson, G. G. 1944. *Tempo and Mode in Eolution*. Columbia University Press, New York.
- Singh, N. D., M. J. Macpherson, J. D. Jensen, y D. A. Petrov. 2007. Similar Levels of X-Linked and Autosomal Nucleotide Variation in African and Non-African Populations of *Drosophila Melanogaster*. *BMC Evolutionary Biology* **7**: 202. doi:10.1186/1471-2148-7-202.
- Singh, N. D., P. F. Arndt, A. G. Clark, y C. F. Aquadro. 2009. Strong Evidence for Lineage and Sequence Specificity of Substitution Rates and Patterns in *Drosophila*. *Molecular Biology and Evolution* **26** (7): 1591–1605. doi:10.1093/molbev/msp071.
- Singh, N. D., J. C. Davis, y D. A. Petrov. 2005. X-Linked Genes Evolve Higher Codon Bias in *Drosophila* and *Caenorhabditis*. *Genetics* **171** (1): 145–155. doi:10.1534/genetics.105.043497.

- 
- Singh, R. S., y L. R. Rhomberg. 1987. A Comprehensive Study of Genic Variation in Natural Populations of *Drosophila Melanogaster*. I. Estimates of Gene Flow from Rare Alleles. *Genetics* **115** (2): 313–22.
- Slotte, T., J. P. Foxe, K. M. Hazzouri, y S. I. Wright. 2010. Genome-Wide Evidence for Efficient Positive and Purifying Selection in *Capsella Grandiflora*, a Plant Species with a Large Effective Population Size. *Molecular Biology and Evolution* **27** (8): 1813–1821. doi:10.1093/molbev/msq062.
- Smith, J. M., y J. Haigh. 1974. The Hitch-Hiking Effect of a Favourable Gene. *Genetical Research* **23** (1): 23–35.
- Smith, N. G. C., M. Brandström, y H. Ellegren. 2004. Evidence for Turnover of Functional Noncoding DNA in Mammalian Genome Evolution. *Genomics* **84** (5): 806–813. doi:10.1016/j.ygeno.2004.07.012.
- Smukowski, C. S., y M. A. F. Noor. 2011. Recombination Rate Variation in Closely Related Species. *Heredity* **107** (6): 496–508. doi:10.1038/hdy.2011.44.
- Sodeland, M., M. Kent, B. J. Hayes, H. Grove, y S. Lien. 2011. Recent and Historical Recombination in the Admixed Norwegian Red Cattle Breed. *BMC Genomics* **12** (1): 33. doi:10.1186/1471-2164-12-33.
- Spencer, C. C. A., P. Deloukas, S. Hunt, J. Mullikin, S. Myers, B. Silverman, P. Donnelly, D. Bentley, y G. McVean. 2006. The Influence of Recombination on Human Genetic Diversity. *PLoS Genetics* **2** (9): 1375–1385. doi:10.1371/journal.pgen.0020148.
- Spies, E. B. 1989. *Genes in Populations*. 2nd editio. John Wiley & Sons, New York.
- Stebbins, G. L. 1950. *Variation and Evolution in Plants*. Columbia University Press, New York.
- Stephan, W., y H. Li. 2007. The Recent Demographic and Adaptive History of *Drosophila Melanogaster*. *Heredity* **98** (2): 65–68. doi:10.1038/sj.hdy.6800901.
- Stephan, W., Y. S. Song, y C. H. Langley. 2006. The Hitchhiking Effect on Linkage Disequilibrium between Linked Neutral Loci. *Genetics* **172** (4): 2647–2663. doi:10.1534/genetics.105.050179.
- Stevison, L. S., y M. A. F. Noor. 2010. Genetic and Evolutionary Correlates of Fine-Scale Recombination Rate Variation in *Drosophila Persimilis*. *Journal of Molecular Evolution* **71** (5-6): 332–345. doi:10.1007/s00239-010-9388-1.
- Stoletzki, N., y A. Eyre-Walker. 2011. Estimation of the Neutrality Index. *Molecular Biology and Evolution* **28** (1): 63–70. doi:10.1093/molbev/msq249.
- Stone, E. A. 2012. Joint Genotyping on the Fly: Identifying Variation among a Sequenced Panel of Inbred Lines. *Genome Research* **22** (5): 966–974. doi:10.1101/gr.129122.111.
- Sturtevant, A. H. 1920. Genetic Studies on *Drosophila Simulans*. I. Introduction. Hybrids with *Drosophila Melanogaster*. *Genetics* **5** (5): 488–500.
- Sung, P, y H. Klein. 2006. Mechanism of Homologous Recombination: Mediators and Helicases Take on Regulatory Functions. *Nature Reviews. Mol Cell Biol* **7** (10): 739–759.
- Tajima, F. 1983. Evolutionary Relationship of DNA Sequences in Finite Populations. *Genetics* **105** (2): 437–60.
- Tajima, F. 1989. Statistical Method for Testing the Neutral Mutation Hypothesis by DNA Polymorphism. *Genetics* **123** (3): 585–95.
- Takano, T. S. 1998. Rate Variation of DNA Sequence Evolution in the *Drosophila* Lineages. *Genetics* **149** (2):959-70.
- Tamura, K., S. Subramanian, y S. Kumar. 2004. Temporal Patterns of Fruit Fly (*Drosophila*) Evolution Revealed by Mutation Clocks. *Molecular Biology and Evolution* **21** (1): 36–44. doi:10.1093/molbev/msg236.
- Tenaillon, M. I., M. C. Sawkins, A. D. Long, R. L. Gaut, J. F. Doebley, y B. S. Gaut. 2001. Patterns of DNA Sequence Polymorphism along Chromosome 1 of Maize (*Zea Mays* Ssp. *Mays* L.). *Proc Natl Acad Sci USA* **98** (16): 9161–9166. doi:10.1073/pnas.151244298.
-

- 
- Tenaillon, M. I., M. C. Sawkins, L. K. Anderson, S. M. Stack, J. Doebley, y B. S. Gaut. 2002. Patterns of Diversity and Recombination along Chromosome 1 of Maize (*Zea Mays* Ssp. *Mays* L.). *Genetics* **162** (3): 1401–1413.
- Tenaillon, M. I., J. U'Ren, O. Tenaillon, y B. S. Gaut. 2004. Selection versus Demography: A Multilocus Investigation of the Domestication Process in Maize. *Molecular Biology and Evolution* **21** (7): 1214–1225. doi:10.1093/molbev/msh102.
- The *C. elegans* Sequencing Consortium. 1998. Genome Sequence of the Nematode *C. Elegans*: A Platform for Investigating Biology. *Science* **282** (5396): 2012–2018. doi:10.1126/science.282.5396.2012.
- The modENCODE Consortium. 2010. Identification of Functional Elements and Regulatory Circuits by *Drosophila* modENCODE. *Science* **330** (6012): 1787–1797. doi:10.1126/science.1198374.
- The 1000 Genomes Project Consortium. 2010. A Map of Human Genome Variation from Population-Scale Sequencing. *Nature* **467**: 1061–1073. doi:10.1038/nature09534.
- Thompson, E. A., S. Deeb, D. Walker, y A. G. Motulsky. 1988. The Detection of Linkage Disequilibrium between Closely Linked Markers: RFLPs at the AI-CIII Apolipoprotein Genes. *American Journal of Human Genetics* **42** (1986): 113–124.
- Thornton, K., y P. Andolfatto. 2006. Approximate Bayesian Inference Reveals Evidence for a Recent, Severe Bottleneck in a Netherlands Population of *Drosophila Melanogaster*. *Genetics* **172** (3): 1607–1619. doi:10.1534/genetics.105.048223.
- Thornton, K., D. Bachtrog, y P. Andolfatto. 2006. X Chromosomes and Autosomes Evolve at Similar Rates in *Drosophila*: No Evidence for Faster-X Protein Evolution. *Genome Research* **16** (4): 498–504. doi:10.1101/gr.4447906.
- Tsai, I. J., A. Burt, y V. Koufopanou. 2010. Conservation of Recombination Hotspots in Yeast. *Proc Natl Acad Sci USA* **107** (17): 7847–7852. doi:10.1073/pnas.0908774107.
- Tupy, J. L., A. M. Bailey, G. Dailey, M. Evans-Holm, C. W. Siebel, S. Misra, S. E. Celniker, y G. M. Rubin. 2005. Identification of Putative Noncoding Polyadenylated Transcripts in *Drosophila Melanogaster*. *Proc Natl Acad Sci USA* **102** (15): 5495–5500. doi:10.1073/pnas.0501422102.
- Vaysse, A., A. Ratnakumar, T. Derrien, E. Axelsson, G. R. Pielberg, S. Sigurdsson, T. Fall, *et al.* 2011. Identification of Genomic Regions Associated with Phenotypic Variation between Dog Breeds Using Selection Mapping. *PLoS Genetics* **7** (10). doi:10.1371/journal.pgen.1002316.
- Venter, J. C., M. D. Adams, E. W. Myers, P. W. Li, R. J. Mural, G. G. Sutton, H. O. Smith, *et al.* 2001. The Sequence of the Human Genome. *Science* **291** (5507): 1304–1351. doi:10.1126/science.1058040.
- Vogl, C., y F. Clemente. 2012. The Allele-Frequency Spectrum in a Decoupled Moran Model with Mutation, Drift, and Directional Selection, Assuming Small Mutation Rates. *Theoretical Population Biology* **81** (3): 197–209. doi:10.1016/j.tpb.2012.01.001.
- Voight, B. F., S. Kudravalli, X. Wen, y J. K. Pritchard. 2006. A Map of Recent Positive Selection in the Human Genome. *PLoS Biology* **4** (3): 0446–0458. doi:10.1371/journal.pbio.0040072.
- Voight, B. F., S. Kudravalli, X. Wen, y J. K. Pritchard. 2006. A Map of Recent Positive Selection in the Human Genome. *PLoS Biology* **4** (3): e72. doi:10.1371/journal.pbio.0040072.
- Wall, J. D., P. Andolfatto, y M. Przeworski. 2002. Testing Models of Selection and Demography in *Drosophila Simulans*. *Genetics* **162** (1): 203–216.
- Walter, M. A., y D. W. Cox. 1991. Nonuniform Linkage Disequilibrium within a 1,500-Kb Region of the Human Immunoglobulin Heavy-Chain Complex. *American Journal of Human Genetics* **49** (5): 917–931.
- Waterston, R. H., K. Lindblad-Toh, E. Birney, J. Rogers, J. F. Abril, P. Agarwal, R. Agarwala, *et al.* 2002. Initial Sequencing and Comparative Analysis of the Mouse Genome. *Nature* **420** (6915): 520–562. doi:10.1038/nature01262.



- 
- Watson, J. D., y F. H. Crick. 1953. Molecular Structure of Nucleic Acids; a Structure for Deoxyribose Nucleic Acid. *Nature* **171** (4356): 737–8.
- Watterson, G. A. 1975. On the Number of Segregating Sites in Genetical Models without Recombination. *Theoretical Population Biology* **7**(2): 256–276. doi:10.1016/0040-5809(75)90020-9.
- Watterson, G. A., y H. A. Guess. 1977. Is the Most Frequent Allele the Oldest? *Theoretical Population Biology* **11** (2): 141–60.
- Webster, M. T., y L. D. Hurst. 2012. Direct and Indirect Consequences of Meiotic Recombination: Implications for Genome Evolution. *Trends in Genetics* **28** (3): 101–109. doi:10.1016/j.tig.2011.11.002.
- Wei, W., L. Ning, Y. Ye, S. Li, H. Zhou, J. Huang, y F. Guo. 2014. SMAL: A Resource of Spontaneous Mutation Accumulation Lines. *Molecular Biology and Evolution* **31** (5): 1302–8. doi:10.1093/molbev/msu073.
- Weir, B S. 1989. Locating the Cystic Fibrosis Gene on the Basis of Linkage Disequilibrium with Markers? *Prog Clin Biol Res* **239**: 81–86.
- Wickham, H. 2009. *ggplot2: Elegant Graphics for Data Analysis*. Springer New York, New York.
- Williamson, S. H., R. Hernandez, A. Fledel-Alon, L. Zhu, R. Nielsen, y C. D. Bustamante. 2005. Simultaneous Inference of Selection and Population Growth from Patterns of Variation in the Human Genome. *Proc Natl Acad Sci USA* **102** (22): 7882–7887. doi:10.1073/pnas.0502300102.
- Williamson, S. 2003. Adaptation in the Env Gene of HIV-1 and Evolutionary Theories of Disease Progression. *Molecular Biology and Evolution* **20** (8): 1318–1325. doi:10.1093/molbev/msg144.
- Williamson, S. H., M. J. Hubisz, A. G. Clark, B. A. Payseur, C. D. Bustamante, y R. Nielsen. 2007. Localizing Recent Adaptive Evolution in the Human Genome. *PLoS Genetics* **3** (6): 0901–0915. doi:10.1371/journal.pgen.0030090.
- Wright, S. 1931. Evolution in Mendelian Populations. *Genetics* **16** (2): 97–159.
- Wright, S. 1950. Genetical Structure of Populations. *Nature* **166** (4215): 247–9.
- Wright, S. 1951. The Genetical Structure of Populations. *Ann Eugen* **15** (4): 323–54.
- Wright, S. I., I. V. Bi, S. G. Schroeder, M. Yamasaki, J. F. Doebley, M. D. McMullen, y B. S. Gaut. 2005. The Effects of Artificial Selection on the Maize Genome. *Science* **308** (5745): 1310–1314. doi:10.1126/science.1107891.
- Xue, Y., Q. Wang, Q. Long, B. L. Ng, H. Swerdlow, J. Burton, C. Skuce, *et al.* 2009. Human Y Chromosome Base-Substitution Mutation Rate Measured by Direct Sequencing in a Deep-Rooting Pedigree. *Current Biology* **19** (17): 1453–1457. doi:10.1016/j.cub.2009.07.032.
- Yang, Z. 1997. PAML: A Program Package for Phylogenetic Analysis by Maximum Likelihood. *Comput Appl Biosci* **13** (5): 555–556.
- Yang, Z. 2007. PAML 4: Phylogenetic Analysis by Maximum Likelihood. *Molecular Biology and Evolution* **24** (8): 1586–1591. doi:10.1093/molbev/msm088.
- Yang, Z., y R. Nielsen. 2008. Mutation-Selection Models of Codon Substitution and Their Use to Estimate Selective Strengths on Codon Usage. *Molecular Biology and Evolution* **25** (3): 568–579. doi:10.1093/molbev/msm284.
- Yeadon P. J., y Catcheside D. E. 1999. Polymorphism around Cog Extends into Adjacent Structural Genes. *Curr Genet* **35** (6): 631–637.
- Zahn, I. 2010. Working with Unbalanced Cell Sizes in Multiple Regression with Categorical Predictors. *Education*: 1–18.
- Zeng, K., y B. Charlesworth. 2009. Estimating Selection Intensity on Synonymous Codon Usage in a Nonequilibrium Population. *Genetics* **183** (2): 651–662. doi:10.1534/genetics.109.101782.

- 
- Zeng, K., y B. Charlesworth. 2010. Studying Patterns of Recent Evolution at Synonymous Sites and Intronic Sites in *Drosophila Melanogaster*. *Journal of Molecular Evolution* **70** (1): 116–128. doi:10.1007/s00239-009-9314-6.
- Zhang, Z., y J. Parsch. 2005. Positive Correlation between Evolutionary Rate and Recombination Rate in *Drosophila* Genes with Male-Biased Expression. *Molecular Biology and Evolution* **22** (10): 1945–1947. doi:10.1093/molbev/msi189.
- Zhou, T., W. Gu, y C. O. Wilke. 2010. Detecting Positive and Purifying Selection at Synonymous Sites in Yeast and Worm. *Molecular Biology and Evolution* **27** (8): 1912–1922. doi:10.1093/molbev/msq077.

---

# Anexos

## 7. Anexos

### 7.1. Tablas Suplementarias

**Tabla suplementaria 1.** *Estimas de diversidad y divergencia en el genoma de D. melanogaster*

$\pi$	Sitios 4 veces degenerados					
	N	$\bar{x}$	P <sub>50</sub>	P <sub>25</sub>	P <sub>75</sub>	$\sigma$
<i>Todos</i>	13.019	0,0112	0,0090	0,0037	0,0165	0,0095
<i>2L</i>	2.400	0,0143	0,0128	0,0060	0,0206	0,0103
<i>Telómero</i>	903	0,0165	0,0154	0,0083	0,0236	0,0108
<i>Centro</i>	910	0,0159	0,0145	0,0082	0,0221	0,0101
<i>Centrómero</i>	587	0,0087	0,0069	0,0033	0,0125	0,0075
<i>2R</i>	2.684	0,0127	0,0106	0,0047	0,0188	0,0101
<i>Telómero</i>	981	0,0152	0,0138	0,0070	0,0220	0,0104
<i>Centro</i>	992	0,0145	0,0133	0,0073	0,0204	0,0099
<i>Centrómero</i>	711	0,0067	0,0049	0,0024	0,0087	0,0068
<i>3L</i>	2.618	0,0123	0,0102	0,0042	0,0183	0,0101
<i>Telómero</i>	924	0,0153	0,0143	0,0075	0,0214	0,0102
<i>Centro</i>	957	0,0147	0,0130	0,0070	0,0203	0,0100
<i>Centrómero</i>	737	0,0054	0,0035	0,0014	0,0075	0,0060
<i>3R</i>	3.270	0,0095	0,0075	0,0030	0,0141	0,0083
<i>Telómero</i>	1.083	0,0132	0,0118	0,0068	0,0178	0,0087
<i>Centro</i>	1.049	0,0116	0,0101	0,0054	0,0162	0,0082
<i>Centrómero</i>	1.138	0,0039	0,0029	0,0009	0,0055	0,0040
<i>X</i>	2.047	0,0071	0,0057	0,0020	0,0107	0,0062
<i>Telómero</i>	672	0,0069	0,0057	0,0019	0,0105	0,0061
<i>Centro</i>	690	0,0077	0,0065	0,0024	0,0113	0,0066
<i>Centrómero</i>	685	0,0065	0,0051	0,0020	0,0098	0,0057
<i>Autosomas</i>	10.972	0,0120	0,0099	0,0042	0,0176	0,0098
<i>Telómero</i>	3.891	0,0150	0,0137	0,0073	0,0211	0,0101
<i>Centro</i>	3.908	0,0141	0,0126	0,0067	0,0196	0,0097
<i>Centrómero</i>	3.173	0,0058	0,0041	0,0015	0,0079	0,0062

$\pi$	Sitios 0 veces degenerados					
	N	$\bar{x}$	P <sub>50</sub>	P <sub>25</sub>	P <sub>75</sub>	$\sigma$
<i>Todos</i>	13.019	0,0014	0,0009	0,0003	0,0019	0,0018
<i>2L</i>	2.400	0,0018	0,0011	0,0004	0,0024	0,0021
<i>Telómero</i>	903	0,0018	0,0011	0,0004	0,0025	0,0022
<i>Centro</i>	910	0,0020	0,0012	0,0005	0,0027	0,0023
<i>Centrómero</i>	587	0,0014	0,0010	0,0004	0,0018	0,0015
<i>2R</i>	2.684	0,0016	0,0010	0,0004	0,0021	0,0021
<i>Telómero</i>	981	0,0018	0,0011	0,0004	0,0022	0,0024
<i>Centro</i>	992	0,0018	0,0011	0,0004	0,0024	0,0022
<i>Centrómero</i>	711	0,0011	0,0008	0,0002	0,0016	0,0012
<i>3L</i>	2.618	0,0015	0,0010	0,0003	0,0021	0,0018
<i>Telómero</i>	924	0,0017	0,0011	0,0004	0,0023	0,0020
<i>Centro</i>	957	0,0018	0,0012	0,0005	0,0025	0,0019
<i>Centrómero</i>	737	0,0010	0,0006	0,0002	0,0013	0,0012
<i>3R</i>	3.270	0,0013	0,0008	0,0003	0,0018	0,0016
<i>Telómero</i>	1.083	0,0016	0,0011	0,0004	0,0022	0,0018
<i>Centro</i>	1.049	0,0015	0,0010	0,0003	0,0020	0,0018
<i>Centrómero</i>	1.138	0,0009	0,0006	0,0002	0,0013	0,0012
<i>X</i>	2.047	0,0008	0,0005	0,0001	0,0012	0,0011
<i>Telómero</i>	672	0,0008	0,0005	0,0001	0,0012	0,0010
<i>Centro</i>	690	0,0008	0,0004	0,0001	0,0011	0,0011
<i>Centrómero</i>	685	0,0009	0,0004	0,0001	0,0012	0,0012
<i>Autosomas</i>	10.972	0,0016	0,0010	0,0003	0,0021	0,0019
<i>Telómero</i>	3.891	0,0017	0,0011	0,0004	0,0023	0,0021
<i>Centro</i>	3.908	0,0018	0,0011	0,0004	0,0024	0,0021
<i>Centrómero</i>	3.173	0,0011	0,0007	0,0002	0,0015	0,0013

$\pi$	Sitios UTR					
	N	$\bar{x}$	P <sub>50</sub>	P <sub>25</sub>	P <sub>75</sub>	$\sigma$
<i>Todos</i>	18.100	0,0045	0,0030	0,0011	0,0060	0,0051
<i>2L</i>	3.285	0,0054	0,0039	0,0016	0,0073	0,0057
<i>Telómero</i>	1.262	0,0061	0,0045	0,0019	0,0081	0,0063
<i>Centro</i>	1.221	0,0056	0,0041	0,0018	0,0075	0,0057
<i>Centrómero</i>	802	0,0040	0,0029	0,0011	0,0055	0,0043
<i>2R</i>	3.725	0,0049	0,0035	0,0014	0,0065	0,0053
<i>Telómero</i>	1.356	0,0058	0,0042	0,0016	0,0077	0,0060
<i>Centro</i>	1.331	0,0053	0,0039	0,0016	0,0068	0,0055
<i>Centrómero</i>	1.038	0,0033	0,0025	0,0009	0,0045	0,0034
<i>3L</i>	3.539	0,0049	0,0034	0,0014	0,0066	0,0053
<i>Telómero</i>	1.300	0,0056	0,0041	0,0019	0,0076	0,0056
<i>Centro</i>	1.225	0,0059	0,0043	0,0020	0,0077	0,0058
<i>Centrómero</i>	1.014	0,0028	0,0019	0,0006	0,0039	0,0032
<i>3R</i>	4.565	0,0042	0,0028	0,0009	0,0055	0,0050
<i>Telómero</i>	1.473	0,0055	0,0039	0,0017	0,0070	0,0059
<i>Centro</i>	1.485	0,0049	0,0034	0,0014	0,0065	0,0053
<i>Centrómero</i>	1.607	0,0024	0,0015	0,0003	0,0033	0,0029
<i>X</i>	2.986	0,0027	0,0017	0,0004	0,0038	0,0033
<i>Telómero</i>	972	0,0028	0,0017	0,0005	0,0038	0,0033
<i>Centro</i>	1.012	0,0027	0,0017	0,0005	0,0037	0,0034
<i>Centrómero</i>	1.002	0,0027	0,0018	0,0004	0,0038	0,0032
<i>Autosomas</i>	15.114	0,0048	0,0033	0,0013	0,0064	0,0053
<i>Telómero</i>	5.391	0,0057	0,0041	0,0018	0,0076	0,0059
<i>Centro</i>	5.262	0,0054	0,0039	0,0017	0,0072	0,0056
<i>Centrómero</i>	4.461	0,0030	0,0020	0,0006	0,0041	0,0034

$\pi$	Sitios intrónicos					
	N	$\bar{x}$	P <sub>50</sub>	P <sub>25</sub>	P <sub>75</sub>	$\sigma$
<i>Todos</i>	6.763	0,0071	0,0050	0,0023	0,0096	0,0068
<i>2L</i>	1.155	0,0092	0,0071	0,0035	0,0126	0,0078
<i>Telómero</i>	422	0,0103	0,0084	0,0041	0,0147	0,0082
<i>Centro</i>	448	0,0104	0,0082	0,0045	0,0139	0,0083
<i>Centrómero</i>	285	0,0057	0,0043	0,0023	0,0077	0,0048
<i>2R</i>	1.408	0,0083	0,0063	0,0031	0,0112	0,0075
<i>Telómero</i>	517	0,0100	0,0075	0,0042	0,0141	0,0082
<i>Centro</i>	501	0,0095	0,0076	0,0042	0,0124	0,0076
<i>Centrómero</i>	390	0,0047	0,0033	0,0019	0,0060	0,0045
<i>3L</i>	1.298	0,0076	0,0057	0,0027	0,0104	0,0069
<i>Telómero</i>	464	0,0090	0,0072	0,0039	0,0127	0,0071
<i>Centro</i>	455	0,0092	0,0070	0,0043	0,0119	0,0075
<i>Centrómero</i>	379	0,0038	0,0026	0,0014	0,0049	0,0039
<i>3R</i>	1.832	0,0059	0,0041	0,0020	0,0081	0,0058
<i>Telómero</i>	580	0,0086	0,0069	0,0037	0,0119	0,0065
<i>Centro</i>	550	0,0072	0,0057	0,0029	0,0096	0,0060
<i>Centrómero</i>	702	0,0028	0,0021	0,0010	0,0039	0,0026
<i>X</i>	1.070	0,0045	0,0031	0,0012	0,0065	0,0044
<i>Telómero</i>	358	0,0044	0,0030	0,0011	0,0067	0,0044
<i>Centro</i>	344	0,0048	0,0032	0,0012	0,0068	0,0049
<i>Centrómero</i>	368	0,0042	0,0031	0,0012	0,0058	0,0039
<i>Autosomas</i>	5.693	0,0076	0,0056	0,0026	0,0103	0,0070
<i>Telómero</i>	1.983	0,0094	0,0075	0,0039	0,0132	0,0075
<i>Centro</i>	1.954	0,0090	0,0070	0,0038	0,0118	0,0074
<i>Centrómero</i>	1.756	0,0039	0,0028	0,0014	0,0050	0,0039

$\pi$	Sitios intergénicos					
	N	$\bar{x}$	P <sub>50</sub>	P <sub>25</sub>	P <sub>75</sub>	$\sigma$
<i>Todos</i>	12.804	0,0056	0,0040	0,0019	0,0074	0,0056
<i>2L</i>	2.496	0,0070	0,0054	0,0028	0,0091	0,0062
<i>Telómero</i>	916	0,0077	0,0058	0,0031	0,0099	0,0069
<i>Centro</i>	919	0,0076	0,0062	0,0033	0,0102	0,0062
<i>Centrómero</i>	661	0,0052	0,0040	0,0020	0,0069	0,0047
<i>2R</i>	1.438	0,0066	0,0048	0,0026	0,0088	0,0059
<i>Telómero</i>	517	0,0073	0,0058	0,0029	0,0095	0,0062
<i>Centro</i>	536	0,0074	0,0057	0,0032	0,0096	0,0065
<i>Centrómero</i>	385	0,0044	0,0033	0,0021	0,0057	0,0038
<i>3L</i>	2.888	0,0062	0,0046	0,0022	0,0082	0,0061
<i>Telómero</i>	996	0,0077	0,0059	0,0031	0,0102	0,0070
<i>Centro</i>	1.070	0,0072	0,0057	0,0031	0,0093	0,0062
<i>Centrómero</i>	822	0,0032	0,0024	0,0012	0,0043	0,0030
<i>3R</i>	3.604	0,0050	0,0035	0,0017	0,0064	0,0050
<i>Telómero</i>	1.124	0,0067	0,0051	0,0028	0,0089	0,0058
<i>Centro</i>	1.104	0,0058	0,0043	0,0021	0,0075	0,0054
<i>Centrómero</i>	1.376	0,0029	0,0022	0,0011	0,0038	0,0029
<i>X</i>	2.378	0,0037	0,0024	0,0010	0,0049	0,0041
<i>Telómero</i>	753	0,0037	0,0023	0,0010	0,0048	0,0041
<i>Centro</i>	810	0,0038	0,0024	0,0009	0,0050	0,0045
<i>Centrómero</i>	815	0,0037	0,0025	0,0011	0,0049	0,0038
<i>Autosomas</i>	10.426	0,0060	0,0044	0,0021	0,0080	0,0058
<i>Telómero</i>	3.553	0,0073	0,0056	0,0029	0,0096	0,0065
<i>Centro</i>	3.629	0,0069	0,0054	0,0029	0,0090	0,0061
<i>Centrómero</i>	3.244	0,0036	0,0027	0,0013	0,0047	0,0036



$\theta$	Sitios 4 veces degenerados					
	N	$\bar{x}$	P <sub>50</sub>	P <sub>25</sub>	P <sub>75</sub>	$\sigma$
<i>Todos</i>	13.019	0,0104	0,0092	0,0051	0,0146	0,0069
<i>2L</i>	2.400	0,0129	0,0122	0,0070	0,0178	0,0073
<i>Telómero</i>	903	0,0147	0,0141	0,0089	0,0198	0,0076
<i>Centro</i>	910	0,0142	0,0137	0,0093	0,0189	0,0068
<i>Centrómero</i>	587	0,0080	0,0068	0,0043	0,0110	0,0052
<i>2R</i>	2.684	0,0113	0,0103	0,0058	0,0158	0,0071
<i>Telómero</i>	981	0,0135	0,0130	0,0079	0,0186	0,0076
<i>Centro</i>	992	0,0126	0,0118	0,0078	0,0165	0,0066
<i>Centrómero</i>	711	0,0066	0,0057	0,0035	0,0085	0,0047
<i>3L</i>	2.618	0,0115	0,0104	0,0056	0,0163	0,0073
<i>Telómero</i>	924	0,0142	0,0139	0,0088	0,0190	0,0072
<i>Centro</i>	957	0,0133	0,0127	0,0082	0,0174	0,0068
<i>Centrómero</i>	737	0,0058	0,0048	0,0029	0,0078	0,0044
<i>3R</i>	3.270	0,0093	0,0083	0,0044	0,0131	0,0062
<i>Telómero</i>	1.083	0,0124	0,0118	0,0080	0,0161	0,0062
<i>Centro</i>	1.049	0,0110	0,0103	0,0069	0,0145	0,0058
<i>Centrómero</i>	1.138	0,0047	0,0042	0,0024	0,0064	0,0033
<i>X</i>	2.047	0,0069	0,0064	0,0036	0,0096	0,0044
<i>Telómero</i>	672	0,0065	0,0059	0,0031	0,0093	0,0046
<i>Centro</i>	690	0,0079	0,0078	0,0045	0,0109	0,0046
<i>Centrómero</i>	685	0,0064	0,0060	0,0034	0,0088	0,0040
<i>Autosomas</i>	10.972	0,0111	0,0100	0,0055	0,0156	0,0071
<i>Telómero</i>	3.891	0,0136	0,0131	0,0084	0,0183	0,0072
<i>Centro</i>	3.908	0,0127	0,0120	0,0079	0,0168	0,0066
<i>Centrómero</i>	3.173	0,0060	0,0051	0,0030	0,0080	0,0044

$\theta$	Sitios 0 veces degenerados					
	N	$\bar{x}$	P <sub>50</sub>	P <sub>25</sub>	P <sub>75</sub>	$\sigma$
<i>Todos</i>	13.019	0,0020	0,0016	0,0008	0,0027	0,0018
<i>2L</i>	2.400	0,0024	0,0019	0,0010	0,0032	0,0020
<i>Telómero</i>	903	0,0024	0,0019	0,0010	0,0032	0,0021
<i>Centro</i>	910	0,0026	0,0020	0,0011	0,0035	0,0022
<i>Centrómero</i>	587	0,0020	0,0016	0,0009	0,0026	0,0015
<i>2R</i>	2.684	0,0022	0,0017	0,0009	0,0028	0,0019
<i>Telómero</i>	981	0,0024	0,0019	0,0010	0,0032	0,0022
<i>Centro</i>	992	0,0023	0,0018	0,0010	0,0030	0,0019
<i>Centrómero</i>	711	0,0017	0,0014	0,0008	0,0023	0,0013
<i>3L</i>	2.618	0,0022	0,0018	0,0009	0,0030	0,0018
<i>Telómero</i>	924	0,0023	0,0020	0,0010	0,0031	0,0019
<i>Centro</i>	957	0,0025	0,0021	0,0011	0,0035	0,0020
<i>Centrómero</i>	737	0,0016	0,0013	0,0007	0,0022	0,0014
<i>3R</i>	3.270	0,0020	0,0015	0,0008	0,0026	0,0017
<i>Telómero</i>	1.083	0,0023	0,0018	0,0010	0,0031	0,0019
<i>Centro</i>	1.049	0,0021	0,0017	0,0009	0,0029	0,0018
<i>Centrómero</i>	1.138	0,0015	0,0012	0,0006	0,0020	0,0012
<i>X</i>	2.047	0,0013	0,0009	0,0004	0,0017	0,0012
<i>Telómero</i>	672	0,0013	0,0009	0,0005	0,0017	0,0012
<i>Centro</i>	690	0,0013	0,0010	0,0004	0,0017	0,0012
<i>Centrómero</i>	685	0,0012	0,0009	0,0004	0,0017	0,0012
<i>Autosomas</i>	10.972	0,0022	0,0017	0,0009	0,0029	0,0019
<i>Telómero</i>	3.891	0,0024	0,0019	0,0010	0,0032	0,0020
<i>Centro</i>	3.908	0,0024	0,0019	0,0010	0,0032	0,0020
<i>Centrómero</i>	3.173	0,0016	0,0013	0,0007	0,0022	0,0014

$\theta$	Sitios UTR					
	N	$\bar{x}$	P <sub>50</sub>	P <sub>25</sub>	P <sub>75</sub>	$\sigma$
<i>Todos</i>	18.100	0,0053	0,0045	0,0027	0,0070	0,0041
<i>2L</i>	3.285	0,0064	0,0055	0,0034	0,0083	0,0045
<i>Telómero</i>	1.262	0,0070	0,0060	0,0036	0,0090	0,0049
<i>Centro</i>	1.221	0,0068	0,0059	0,0038	0,0089	0,0044
<i>Centrómero</i>	802	0,0049	0,0042	0,0026	0,0063	0,0034
<i>2R</i>	3.725	0,0057	0,0048	0,0029	0,0073	0,0041
<i>Telómero</i>	1.356	0,0064	0,0055	0,0033	0,0085	0,0046
<i>Centro</i>	1.331	0,0060	0,0052	0,0032	0,0079	0,0041
<i>Centrómero</i>	1.038	0,0042	0,0037	0,0023	0,0056	0,0030
<i>3L</i>	3.539	0,0059	0,0051	0,0030	0,0078	0,0042
<i>Telómero</i>	1.300	0,0065	0,0056	0,0036	0,0086	0,0043
<i>Centro</i>	1.225	0,0069	0,0060	0,0039	0,0086	0,0044
<i>Centrómero</i>	1.014	0,0039	0,0034	0,0020	0,0053	0,0028
<i>3R</i>	4.565	0,0051	0,0043	0,0026	0,0066	0,0039
<i>Telómero</i>	1.473	0,0063	0,0054	0,0034	0,0082	0,0043
<i>Centro</i>	1.485	0,0058	0,0050	0,0032	0,0072	0,0041
<i>Centrómero</i>	1.607	0,0033	0,0030	0,0016	0,0045	0,0026
<i>X</i>	2.986	0,0035	0,0030	0,0017	0,0047	0,0027
<i>Telómero</i>	972	0,0035	0,0029	0,0017	0,0047	0,0028
<i>Centro</i>	1.012	0,0037	0,0032	0,0019	0,0050	0,0028
<i>Centrómero</i>	1.002	0,0033	0,0029	0,0016	0,0045	0,0025
<i>Autosomas</i>	15.114	0,0057	0,0048	0,0029	0,0074	0,0042
<i>Telómero</i>	5.391	0,0065	0,0056	0,0035	0,0085	0,0045
<i>Centro</i>	5.262	0,0063	0,0055	0,0035	0,0082	0,0043
<i>Centrómero</i>	4.461	0,0039	0,0034	0,0020	0,0053	0,0029

$\theta$	Sitios intrónicos					
	N	$\bar{x}$	P <sub>50</sub>	P <sub>25</sub>	P <sub>75</sub>	$\sigma$
<i>Todos</i>	6.763	0,0072	0,0061	0,0035	0,0097	0,0050
<i>2L</i>	1.155	0,0091	0,0080	0,0049	0,0120	0,0057
<i>Telómero</i>	422	0,0101	0,0091	0,0057	0,0133	0,0060
<i>Centro</i>	448	0,0103	0,0088	0,0062	0,0129	0,0058
<i>Centrómero</i>	285	0,0059	0,0050	0,0032	0,0076	0,0037
<i>2R</i>	1.408	0,0079	0,0067	0,0041	0,0107	0,0052
<i>Telómero</i>	517	0,0094	0,0083	0,0050	0,0126	0,0058
<i>Centro</i>	501	0,0088	0,0077	0,0052	0,0114	0,0049
<i>Centrómero</i>	390	0,0050	0,0041	0,0028	0,0063	0,0032
<i>3L</i>	1.298	0,0077	0,0068	0,0039	0,0105	0,0050
<i>Telómero</i>	464	0,0092	0,0083	0,0056	0,0123	0,0051
<i>Centro</i>	455	0,0090	0,0081	0,0054	0,0115	0,0049
<i>Centrómero</i>	379	0,0044	0,0035	0,0023	0,0058	0,0029
<i>3R</i>	1.832	0,0064	0,0053	0,0031	0,0086	0,0045
<i>Telómero</i>	580	0,0085	0,0077	0,0050	0,0113	0,0047
<i>Centro</i>	550	0,0076	0,0067	0,0045	0,0097	0,0045
<i>Centrómero</i>	702	0,0036	0,0031	0,0020	0,0046	0,0024
<i>X</i>	1.070	0,0048	0,0042	0,0026	0,0063	0,0032
<i>Telómero</i>	358	0,0046	0,0040	0,0022	0,0062	0,0032
<i>Centro</i>	344	0,0053	0,0046	0,0027	0,0070	0,0035
<i>Centrómero</i>	368	0,0045	0,0040	0,0028	0,0058	0,0027
<i>Autosomas</i>	5.693	0,0076	0,0065	0,0038	0,0102	0,0051
<i>Telómero</i>	1.983	0,0093	0,0082	0,0053	0,0123	0,0054
<i>Centro</i>	1.954	0,0088	0,0078	0,0052	0,0114	0,0051
<i>Centrómero</i>	1.756	0,0044	0,0037	0,0024	0,0057	0,0030

$\theta$	Sitios intergénicos					
	N	$\bar{x}$	P <sub>50</sub>	P <sub>25</sub>	P <sub>75</sub>	$\sigma$
<i>Todos</i>	12.804	0,0063	0,0053	0,0032	0,0083	0,0044
<i>2L</i>	2.496	0,0077	0,0068	0,0042	0,0100	0,0047
<i>Telómero</i>	916	0,0082	0,0072	0,0045	0,0106	0,0051
<i>Centro</i>	919	0,0085	0,0078	0,0053	0,0110	0,0045
<i>Centrómero</i>	661	0,0058	0,0049	0,0031	0,0076	0,0039
<i>2R</i>	1.438	0,0071	0,0062	0,0040	0,0092	0,0046
<i>Telómero</i>	517	0,0080	0,0070	0,0044	0,0102	0,0048
<i>Centro</i>	536	0,0078	0,0066	0,0045	0,0099	0,0048
<i>Centrómero</i>	385	0,0051	0,0045	0,0030	0,0065	0,0030
<i>3L</i>	2.888	0,0070	0,0060	0,0037	0,0091	0,0047
<i>Telómero</i>	996	0,0083	0,0072	0,0049	0,0105	0,0052
<i>Centro</i>	1.070	0,0079	0,0068	0,0049	0,0099	0,0046
<i>Centrómero</i>	822	0,0042	0,0036	0,0024	0,0055	0,0027
<i>3R</i>	3.604	0,0058	0,0049	0,0030	0,0075	0,0041
<i>Telómero</i>	1.124	0,0074	0,0064	0,0043	0,0093	0,0044
<i>Centro</i>	1.104	0,0067	0,0057	0,0039	0,0084	0,0041
<i>Centrómero</i>	1.376	0,0039	0,0034	0,0021	0,0051	0,0026
<i>X</i>	2.378	0,0044	0,0038	0,0023	0,0057	0,0031
<i>Telómero</i>	753	0,0042	0,0034	0,0020	0,0054	0,0031
<i>Centro</i>	810	0,0049	0,0042	0,0024	0,0064	0,0034
<i>Centrómero</i>	815	0,0042	0,0037	0,0023	0,0054	0,0027
<i>Autosomas</i>	10.426	0,0068	0,0058	0,0036	0,0088	0,0045
<i>Telómero</i>	3.553	0,0079	0,0070	0,0045	0,0101	0,0049
<i>Centro</i>	3.629	0,0077	0,0067	0,0045	0,0098	0,0045
<i>Centrómero</i>	3.244	0,0045	0,0038	0,0024	0,0058	0,0031

<i>k</i>	Sitios 4 veces degenerados					
	<b>N</b>	$\bar{x}$	<b>P<sub>50</sub></b>	<b>P<sub>25</sub></b>	<b>P<sub>75</sub></b>	<b><math>\sigma</math></b>
<i>Todos</i>	11.312	0,2463	0,2413	0,1885	0,2971	0,0917
<i>2L</i>	2.074	0,2506	0,2471	0,1971	0,2988	0,0876
<i>Telómero</i>	781	0,2469	0,2380	0,1938	0,2928	0,0906
<i>Centro</i>	790	0,2498	0,2494	0,1955	0,3011	0,0855
<i>Centrómero</i>	503	0,2576	0,2573	0,2044	0,3033	0,0858
<i>2R</i>	2.340	0,2467	0,2416	0,1893	0,2974	0,0933
<i>Telómero</i>	872	0,2446	0,2396	0,1857	0,2980	0,0920
<i>Centro</i>	858	0,2475	0,2430	0,1851	0,2974	0,1021
<i>Centrómero</i>	610	0,2487	0,2434	0,1986	0,2964	0,0815
<i>3L</i>	2.301	0,2428	0,2388	0,1859	0,2943	0,0909
<i>Telómero</i>	924	0,2412	0,2354	0,1829	0,2946	0,0932
<i>Centro</i>	828	0,2432	0,2363	0,1820	0,2982	0,0983
<i>Centrómero</i>	638	0,2492	0,2454	0,1962	0,2927	0,0891
<i>3R</i>	2.858	0,2450	0,2423	0,1885	0,2963	0,0913
<i>Telómero</i>	960	0,2423	0,2362	0,1834	0,2975	0,0963
<i>Centro</i>	904	0,2427	0,2428	0,1830	0,2911	0,0934
<i>Centrómero</i>	994	0,2496	0,2492	0,1974	0,3004	0,0840
<i>X</i>	1.739	0,2478	0,2384	0,1859	0,2982	0,0959
<i>Telómero</i>	569	0,2632	0,2534	0,1988	0,3176	0,0962
<i>Centro</i>	601	0,2487	0,2396	0,1885	0,2947	0,0995
<i>Centrómero</i>	569	0,2313	0,2217	0,1722	0,2787	0,0891
<i>Autosomas</i>	9.573	0,2461	0,2421	0,1895	0,2969	0,0909
<i>Telómero</i>	3.448	0,2427	0,2368	0,1860	0,2950	0,0911
<i>Centro</i>	3.380	0,2457	0,2430	0,1855	0,2971	0,0952
<i>Centrómero</i>	2.745	0,2508	0,2490	0,1989	0,2983	0,0850

<i>k</i>	Sitios 0 veces degenerados					
	<b>N</b>	$\bar{x}$	<b>P<sub>50</sub></b>	<b>P<sub>25</sub></b>	<b>P<sub>75</sub></b>	<b><math>\sigma</math></b>
<i>Todos</i>	11.312	0,0456	0,0267	0,0121	0,0558	0,0580
<i>2L</i>	2.074	0,0458	0,0259	0,0122	0,0558	0,0579
<i>Telómero</i>	781	0,0425	0,0226	0,0105	0,0498	0,0565
<i>Centro</i>	790	0,0461	0,0259	0,0119	0,0557	0,0597
<i>Centrómero</i>	503	0,0504	0,0329	0,0152	0,0623	0,0571
<i>2R</i>	2.340	0,0457	0,0269	0,0121	0,0548	0,0599
<i>Telómero</i>	872	0,0454	0,0270	0,0127	0,0563	0,0554
<i>Centro</i>	858	0,0464	0,0244	0,0110	0,0532	0,0673
<i>Centrómero</i>	610	0,0450	0,0307	0,0134	0,0549	0,0549
<i>3L</i>	2.301	0,0423	0,0250	0,0117	0,0510	0,0533
<i>Telómero</i>	835	0,0377	0,0224	0,0109	0,0465	0,0456
<i>Centro</i>	828	0,0456	0,0265	0,0112	0,0555	0,0591
<i>Centrómero</i>	638	0,0441	0,0280	0,0132	0,0519	0,0543
<i>3R</i>	2.858	0,0428	0,0255	0,0113	0,0543	0,0526
<i>Telómero</i>	960	0,0414	0,0240	0,0099	0,0516	0,0541
<i>Centro</i>	904	0,0426	0,0248	0,0112	0,0576	0,0493
<i>Centrómero</i>	994	0,0445	0,0271	0,0131	0,0547	0,0540
<i>X</i>	1.739	0,0539	0,0322	0,0138	0,0657	0,0681
<i>Telómero</i>	569	0,0549	0,0345	0,0161	0,0662	0,0627
<i>Centro</i>	601	0,0547	0,0319	0,0132	0,0654	0,0743
<i>Centrómero</i>	569	0,0520	0,0302	0,0114	0,0655	0,0665
<i>Autosomas</i>	9.573	0,0440	0,0258	0,0118	0,0540	0,0558
<i>Telómero</i>	3.448	0,0418	0,0239	0,0110	0,0503	0,0531
<i>Centro</i>	3.380	0,0451	0,0255	0,0113	0,0554	0,0591
<i>Centrómero</i>	2.745	0,0456	0,0290	0,0134	0,0563	0,0549

<i>k</i>	Sitios UTR					
	<b>N</b>	$\bar{x}$	<b>P<sub>50</sub></b>	<b>P<sub>25</sub></b>	<b>P<sub>75</sub></b>	<b><math>\sigma</math></b>
<i>Todos</i>	14.502	0,1198	0,0979	0,0615	0,1540	0,0852
<i>2L</i>	2.674	0,1219	0,1001	0,0629	0,1585	0,0853
<i>Telómero</i>	1.060	0,1185	0,0961	0,0613	0,1519	0,0835
<i>Centro</i>	1.011	0,1152	0,0928	0,0548	0,1536	0,0854
<i>Centrómero</i>	603	0,1393	0,1166	0,0816	0,1764	0,0858
<i>2R</i>	3.021	0,1139	0,0906	0,0560	0,1451	0,0845
<i>Telómero</i>	1.126	0,1101	0,0860	0,0525	0,1394	0,0850
<i>Centro</i>	1.095	0,1072	0,0847	0,0528	0,1381	0,0811
<i>Centrómero</i>	800	0,1282	0,1064	0,0710	0,1598	0,0868
<i>3L</i>	2.923	0,1141	0,0920	0,0594	0,1444	0,0817
<i>Telómero</i>	1.097	0,1128	0,0910	0,0570	0,1399	0,0843
<i>Centro</i>	1.013	0,1130	0,0895	0,0587	0,1451	0,0809
<i>Centrómero</i>	813	0,1171	0,0967	0,0630	0,1482	0,0791
<i>3R</i>	3.631	0,1202	0,0974	0,0608	0,1556	0,0865
<i>Telómero</i>	1.196	0,1095	0,0875	0,0551	0,1394	0,0816
<i>Centro</i>	1.183	0,1211	0,0960	0,0607	0,1556	0,0882
<i>Centrómero</i>	1.252	0,1296	0,1112	0,0691	0,1668	0,0882
<i>X</i>	2.253	0,1320	0,1110	0,0729	0,1690	0,0867
<i>Telómero</i>	716	0,1405	0,1210	0,0793	0,1774	0,0876
<i>Centro</i>	777	0,1309	0,1083	0,0727	0,1664	0,0869
<i>Centrómero</i>	760	0,1251	0,1042	0,0684	0,1620	0,0852
<i>Autosomas</i>	12.249	0,1176	0,0956	0,0598	0,1511	0,0847
<i>Telómero</i>	4.479	0,1126	0,0898	0,0564	0,1426	0,0836
<i>Centro</i>	4.302	0,1143	0,0907	0,0566	0,1486	0,0842
<i>Centrómero</i>	3.468	0,1280	0,1073	0,0698	0,1631	0,0857



<i>k</i>	Sitios intrónicos					
	<b>N</b>	$\bar{x}$	<b>P<sub>50</sub></b>	<b>P<sub>25</sub></b>	<b>P<sub>75</sub></b>	<b><math>\sigma</math></b>
<i>Todos</i>	4.943	0,2207	0,2134	0,1607	0,2708	0,0805
<i>2L</i>	870	0,2323	0,2260	0,1734	0,2798	0,0845
<i>Telómero</i>	329	0,2224	0,2109	0,1575	0,2742	0,0834
<i>Centro</i>	338	0,2375	0,2291	0,1871	0,2835	0,0817
<i>Centrómero</i>	203	0,2397	0,2363	0,1794	0,2790	0,0897
<i>2R</i>	1.078	0,2120	0,2058	0,1512	0,2630	0,0776
<i>Telómero</i>	410	0,2076	0,2009	0,1447	0,2630	0,0790
<i>Centro</i>	397	0,2146	0,2088	0,1518	0,2641	0,0801
<i>Centrómero</i>	271	0,2149	0,2125	0,1629	0,2608	0,0716
<i>3L</i>	990	0,2186	0,2106	0,1580	0,2712	0,0821
<i>Telómero</i>	371	0,2075	0,1955	0,1492	0,2575	0,0794
<i>Centro</i>	353	0,2206	0,2137	0,1550	0,2767	0,0843
<i>Centrómero</i>	266	0,2315	0,2258	0,1705	0,2819	0,0811
<i>3R</i>	1.384	0,2210	0,2140	0,1594	0,2714	0,0788
<i>Telómero</i>	461	0,2110	0,2014	0,1486	0,2687	0,0772
<i>Centro</i>	420	0,2299	0,2238	0,1705	0,2757	0,0799
<i>Centrómero</i>	503	0,2228	0,2125	0,1658	0,2667	0,0785
<i>X</i>	621	0,2220	0,2102	0,1661	0,2680	0,0788
<i>Telómero</i>	207	0,2110	0,2024	0,1592	0,2581	0,0712
<i>Centro</i>	220	0,2336	0,2257	0,1734	0,2829	0,0850
<i>Centrómero</i>	194	0,2207	0,2078	0,1663	0,2642	0,0779
<i>Autosomas</i>	4.322	0,2205	0,2136	0,1595	0,2712	0,0807
<i>Telómero</i>	1.571	0,2117	0,2030	0,1512	0,2678	0,0797
<i>Centro</i>	1.508	0,2254	0,2203	0,1657	0,2748	0,0818
<i>Centrómero</i>	1.243	0,2257	0,2196	0,1686	0,2702	0,0799

<i>k</i>	Sitios intergénicos					
	<b>N</b>	$\bar{x}$	<b>P<sub>50</sub></b>	<b>P<sub>25</sub></b>	<b>P<sub>75</sub></b>	<b><math>\sigma</math></b>
<i>Todos</i>	7.856	0,1908	0,1648	0,1090	0,2587	0,1085
<i>2L</i>	1.525	0,1982	0,1757	0,1161	0,2657	0,1073
<i>Telómero</i>	620	0,1827	0,1595	0,1068	0,2500	0,0963
<i>Centro</i>	587	0,2051	0,1885	0,1202	0,2747	0,1097
<i>Centrómero</i>	318	0,2157	0,1922	0,1294	0,2864	0,1188
<i>2R</i>	874	0,1889	0,1594	0,1067	0,2546	0,1115
<i>Telómero</i>	336	0,1793	0,1494	0,1057	0,2354	0,1007
<i>Centro</i>	339	0,1839	0,1558	0,1048	0,2503	0,1082
<i>Centrómero</i>	199	0,2138	0,1762	0,1150	0,2995	0,1298
<i>3L</i>	1.866	0,1851	0,1568	0,1025	0,2562	0,1102
<i>Telómero</i>	688	0,1697	0,1434	0,0948	0,2363	0,1000
<i>Centro</i>	691	0,1883	0,1595	0,1017	0,2609	0,1155
<i>Centrómero</i>	487	0,2022	0,1715	0,1212	0,2744	0,1137
<i>3R</i>	2.267	0,1866	0,1574	0,1030	0,2547	0,1113
<i>Telómero</i>	776	0,1746	0,1478	0,0979	0,2392	0,1052
<i>Centro</i>	703	0,1874	0,1624	0,1040	0,2611	0,1081
<i>Centrómero</i>	788	0,1977	0,1658	0,1106	0,2660	0,1185
<i>X</i>	1.324	0,1985	0,1768	0,1234	0,2630	0,0996
<i>Telómero</i>	421	0,1818	0,1561	0,1116	0,2428	0,0924
<i>Centro</i>	484	0,2037	0,1805	0,1264	0,2680	0,1032
<i>Centrómero</i>	419	0,2093	0,1915	0,1316	0,2744	0,1005
<i>Autosomas</i>	6.532	0,1892	0,1614	0,1064	0,2580	0,1102
<i>Telómero</i>	2.420	0,1759	0,1505	0,0997	0,2397	0,1009
<i>Centro</i>	2.320	0,1916	0,1668	0,1072	0,2627	0,1110
<i>Centrómero</i>	1.792	0,2039	0,1730	0,1153	0,2738	0,1187

Estimas de diversidad nucleotídica ( $\pi$ ,  $\theta$ ) y de divergencia ( $k$ ) para cada brazo cromosómico y clase funcional. También se muestran los promedios de las distintas regiones cromosómicas así como el número de datos utilizados para calcular la media, mediana y desviación estándar.

**Tabla suplementaria 2. Mapa de bloques de selección ligada**

<i>IMb</i>	<i>ρ</i>			
	LSB		NLSB	
	Inicio	Fin	Inicio	Fin
<i>2L</i>	1	1.000.000	1.000.001	2.000.000
	2.000.001	3.000.000	3.000.001	5.000.000
	5.000.001	6.000.000	6.000.001	7.000.000
	7.000.001	10.000.000	10.000.001	11.000.000
	11.000.001	23.000.000		
<i>2R</i>	1	6.000.000	6.000.001	8.000.000
	8.000.001	9.000.000	9.000.001	11.000.000
	11.000.001	12.000.000	12.000.001	20.000.000
	20.000.001	21.000.000		
<i>3L</i>	1	2.000.000	2.000.001	3.000.000
	3.000.001	12.000.000	12.000.001	13.000.000
	13.000.001	24.000.000		
<i>3R</i>	1	9.000.000	9.000.001	11.000.000
	11.000.001	13.000.000	13.000.001	15.000.000
	15.000.001	17.000.000	17.000.001	24.000.000
	24.000.001	25.000.000	25.000.001	27.000.000
<i>X</i>	1	2.000.000	2.000.001	21.000.000
	21.000.001	22.000.000		

<i>100kb</i>	$\rho$			
	LSB		NLSB	
	Inicio	Fin	Inicio	Fin
<i>2L</i>	1	800.000	800.001	1.000.000
	1.000.001	1.200.000	1.200.001	1.300.000
	1.300.001	1.400.000	1.400.001	1.500.000
	1.500.001	1.600.000	1.600.001	1.700.000
	1.700.001	1.900.000	1.900.001	2.000.000
	2.000.001	2.600.000	2.600.001	2.700.000
	2.700.001	3.100.000	3.100.001	3.200.000
	3.200.001	3.400.000	3.400.001	4.100.000
	4.100.001	4.200.000	4.200.001	4.300.000
	4.300.001	4.700.000	4.700.001	4.800.000
	4.800.001	5.000.000	5.000.001	5.100.000
	5.100.001	5.400.000	5.400.001	5.700.000
	5.700.001	6.000.000	6.000.001	6.100.000
	6.100.001	6.200.000	6.200.001	6.400.000
	6.400.001	6.500.000	6.500.001	6.600.000
	6.600.001	6.700.000	6.700.001	6.900.000
	6.900.001	7.000.000	7.000.001	7.300.000
	7.300.001	8.500.000	8.500.001	8.700.000
	8.700.001	9.100.000	9.100.001	9.200.000
	9.200.001	9.500.000	9.500.001	9.700.000
	9.700.001	9.800.000	9.800.001	10.800.000
	10.800.001	10.900.000	10.900.001	11.000.000
	11.000.001	11.300.000	11.300.001	11.400.000
	11.400.001	12.100.000	12.100.001	12.600.000
	12.600.001	13.900.000	13.900.001	14.100.000
	14.100.001	14.200.000	14.200.001	14.400.000
	14.400.001	15.100.000	15.100.001	15.500.000
	15.500.001	15.700.000	15.700.001	15.800.000
15.800.001	16.300.000	16.300.001	16.600.000	
16.600.001	23.100.000			

<i>100kb</i>	$\rho$			
	LSB		NLSB	
	Inicio	Fin	Inicio	Fin
<i>2R</i>	1	6.200.000	6.200.001	6.400.000
	6.400.001	6.500.000	6.500.001	7.100.000
	7.100.001	7.300.000	7.300.001	7.600.000
	7.600.001	8.600.000	8.600.001	8.700.000
	8.700.001	8.900.000	8.900.001	9.000.000
	9.000.001	9.200.000	9.200.001	9.800.000
	9.800.001	9.900.000	9.900.001	10.000.000
	10.000.001	10.500.000	10.500.001	10.700.000
	10.700.001	10.800.000	10.800.001	11.000.000
	11.000.001	11.600.000	11.600.001	11.800.000
	11.800.001	11.900.000	11.900.001	12.500.000
	12.500.001	12.600.000	12.600.001	12.700.000
	12.700.001	13.500.000	13.500.001	13.600.000
	13.600.001	13.700.000	13.700.001	14.000.000
	14.000.001	14.100.000	14.100.001	14.300.000
	14.300.001	14.400.000	14.400.001	15.300.000
	15.300.001	15.400.000	15.400.001	15.500.000
	15.500.001	15.900.000	15.900.001	16.500.000
	16.500.001	16.600.000	16.600.001	17.200.000
	17.200.001	17.300.000	17.300.001	18.100.000
	18.100.001	18.300.000	18.300.001	18.600.000
	18.600.001	18.800.000	18.800.001	19.200.000
19.200.001	19.300.000	19.300.001	19.800.000	
19.800.001	21.200.000			

<i>100kb</i>	$\rho$			
	LSB		NLSB	
	Inicio	Fin	Inicio	Fin
<i>3L</i>	1	1.000.000	1.000.001	1.100.000
	1.100.001	1.400.000	1.400.001	1.500.000
	1.500.001	2.300.000	2.300.001	2.400.000
	2.400.001	2.500.000	2.500.001	2.600.000
	2.600.001	3.500.000	3.500.001	3.600.000
	3.600.001	3.700.000	3.700.001	3.800.000
	3.800.001	4.000.000	4.000.001	4.100.000
	4.100.001	4.600.000	4.600.001	4.800.000
	4.800.001	5.300.000	5.300.001	5.500.000
	5.500.001	6.400.000	6.400.001	6.500.000
	6.500.001	7.200.000	7.200.001	7.300.000
	7.300.001	8.000.000	8.000.001	8.100.000
	8.100.001	9.200.000	9.200.001	9.300.000
	9.300.001	10.400.000	10.400.001	10.500.000
	10.500.001	10.700.000	10.700.001	10.800.000
	10.800.001	10.900.000	10.900.001	11.000.000
	11.000.001	11.300.000	11.300.001	11.500.000
	11.500.001	12.000.000	12.000.001	12.100.000
	12.100.001	12.400.000	12.400.001	12.900.000
	12.900.001	13.100.000	13.100.001	13.200.000
13.200.001	18.200.000	18.200.001	18.300.000	
18.300.001	24.600.000			

<i>100kb</i>	$\rho$			
	LSB		NLSB	
	Inicio	Fin	Inicio	Fin
<i>3R</i>	1	9.500.000	9.500.001	10.000.000
	10.000.001	10.100.000	10.100.001	10.300.000
	10.300.001	10.500.000	10.500.001	10.600.000
	10.600.001	10.800.000	10.800.001	10.900.000
	10.900.001	11.200.000	11.200.001	11.300.000
	11.300.001	13.600.000	13.600.001	14.500.000
	14.500.001	14.600.000	14.600.001	15.300.000
	15.300.001	17.000.000	17.000.001	17.100.000
	17.100.001	17.600.000	17.600.001	18.100.000
	18.100.001	18.200.000	18.200.001	19.300.000
	19.300.001	19.400.000	19.400.001	20.100.000
	20.100.001	20.200.000	20.200.001	21.000.000
	21.000.001	21.300.000	21.300.001	21.500.000
	21.500.001	21.600.000	21.600.001	21.700.000
	21.700.001	22.600.000	22.600.001	22.800.000
	22.800.001	23.200.000	23.200.001	23.500.000
	23.500.001	23.800.000	23.800.001	23.900.000
	23.900.001	25.000.000	25.000.001	25.700.000
	25.700.001	26.000.000	26.000.001	26.200.000
	26.200.001	26.600.000	26.600.001	26.700.000
26.700.001	28.000.000			

<i>100kb</i>	<i>ρ</i>			
	LSB		NLSB	
	Inicio	Fin	Inicio	Fin
<i>X</i>	1	2.200.000	2.200.001	3.500.000
	3.500.001	3.700.000	3.700.001	7.400.000
	7.400.001	7.700.000	7.700.001	9.500.000
	9.500.001	9.600.000	9.600.001	10.300.000
	10.300.001	10.400.000	10.400.001	14.200.000
	14.200.001	14.300.000	14.300.001	14.800.000
	14.800.001	15.100.000	15.100.001	17.000.000
	17.000.001	17.200.000	17.200.001	17.300.000
	17.300.001	17.400.000	17.400.001	17.900.000
	17.900.001	18.300.000	18.300.001	18.500.000
	18.500.001	18.600.000	18.600.001	19.500.000
	19.500.001	19.600.000	19.600.001	20.100.000
	20.100.001	20.200.000	20.200.001	20.600.000
	20.600.001	20.900.000	20.900.001	21.100.000
	21.100.001	22.500.000		



<i>10kb</i>	$\rho$			
	LSB		NLSB	
	Inicio	Fin	Inicio	Fin
<i>2L</i>	1	500.000	500.001	520.000
	520.001	580.000	580.001	590.000
	590.001	610.000	610.001	620.000
	620.001	790.000	790.001	810.000
	810.001	830.000	830.001	860.000
	860.001	880.000	880.001	920.000
	920.001	930.000	930.001	980.000
	980.001	1.020.000	1.020.001	1.040.000
	1.040.001	1.080.000	1.080.001	1.100.000
	1.100.001	1.130.000	1.130.001	1.150.000
	1.150.001	1.180.000	1.180.001	1.220.000
	1.220.001	1.230.000	1.230.001	1.330.000
	1.330.001	1.340.000	1.340.001	1.350.000
	1.350.001	1.430.000	1.430.001	1.440.000
	1.440.001	1.450.000	1.450.001	1.480.000
	1.480.001	1.610.000	1.610.001	1.640.000
	1.640.001	1.650.000	1.650.001	1.730.000
	1.730.001	1.820.000	1.820.001	1.860.000
	1.860.001	1.870.000	1.870.001	1.890.000
	1.890.001	1.900.000	1.900.001	1.930.000
	1.930.001	1.950.000	1.950.001	1.970.000
	1.970.001	1.990.000	1.990.001	2.000.000
	2.000.001	2.020.000	2.020.001	2.030.000
	2.030.001	2.060.000	2.060.001	2.070.000
	2.070.001	2.090.000	2.090.001	2.110.000
	2.110.001	2.120.000	2.120.001	2.130.000
	2.130.001	2.160.000	2.160.001	2.170.000
	2.170.001	2.220.000	2.220.001	2.230.000
	2.230.001	2.390.000	2.390.001	2.400.000
	2.400.001	2.500.000	2.500.001	2.520.000
2.520.001	2.540.000	2.540.001	2.550.000	
2.550.001	2.590.000	2.590.001	2.600.000	
2.600.001	2.640.000	2.640.001	2.700.000	
2.700.001	2.710.000	2.710.001	2.740.000	
2.740.001	2.790.000	2.790.001	2.800.000	

<i>10kb</i>	$\rho$			
	LSB		NLSB	
	Inicio	Fin	Inicio	Fin
<i>2L</i>	2.800.001	2.830.000	2.830.001	2.850.000
	2.850.001	2.860.000	2.860.001	2.880.000
	2.880.001	2.910.000	2.910.001	2.920.000
	2.920.001	3.080.000	3.080.001	3.160.000
	3.160.001	3.180.000	3.180.001	3.250.000
	3.250.001	3.260.000	3.260.001	3.280.000
	3.280.001	3.310.000	3.310.001	3.320.000
	3.320.001	3.380.000	3.380.001	3.450.000
	3.450.001	3.460.000	3.460.001	3.510.000
	3.510.001	3.550.000	3.550.001	3.560.000
	3.560.001	3.570.000	3.570.001	3.610.000
	3.610.001	3.620.000	3.620.001	3.650.000
	3.650.001	3.660.000	3.660.001	3.710.000
	3.710.001	3.720.000	3.720.001	3.800.000
	3.800.001	3.810.000	3.810.001	3.910.000
	3.910.001	3.930.000	3.930.001	3.970.000
	3.970.001	3.980.000	3.980.001	4.060.000
	4.060.001	4.070.000	4.070.001	4.150.000
	4.150.001	4.160.000	4.160.001	4.170.000
	4.170.001	4.200.000	4.200.001	4.340.000
	4.340.001	4.370.000	4.370.001	4.380.000
	4.380.001	4.420.000	4.420.001	4.440.000
	4.440.001	4.650.000	4.650.001	4.660.000
	4.660.001	4.670.000	4.670.001	4.720.000
	4.720.001	4.730.000	4.730.001	4.770.000
	4.770.001	4.780.000	4.780.001	4.810.000
	4.810.001	4.850.000	4.850.001	4.880.000
	4.880.001	4.890.000	4.890.001	4.910.000
	4.910.001	4.920.000	4.920.001	4.940.000
	4.940.001	4.960.000	4.960.001	4.970.000
4.970.001	5.010.000	5.010.001	5.030.000	
5.030.001	5.060.000	5.060.001	5.070.000	
5.070.001	5.080.000	5.080.001	5.090.000	
5.090.001	5.100.000	5.100.001	5.120.000	
5.120.001	5.230.000	5.230.001	5.240.000	

<i>10kb</i>	$\rho$			
	LSB		NLSB	
	Inicio	Fin	Inicio	Fin
<i>2L</i>	5.240.001	5.250.000	5.250.001	5.270.000
	5.270.001	5.370.000	5.370.001	5.400.000
	5.400.001	5.410.000	5.410.001	5.460.000
	5.460.001	5.470.000	5.470.001	5.520.000
	5.520.001	5.530.000	5.530.001	5.540.000
	5.540.001	5.560.000	5.560.001	5.580.000
	5.580.001	5.590.000	5.590.001	5.660.000
	5.660.001	5.680.000	5.680.001	5.720.000
	5.720.001	5.790.000	5.790.001	5.800.000
	5.800.001	5.810.000	5.810.001	5.820.000
	5.820.001	5.840.000	5.840.001	5.900.000
	5.900.001	5.970.000	5.970.001	5.980.000
	5.980.001	6.010.000	6.010.001	6.040.000
	6.040.001	6.090.000	6.090.001	6.160.000
	6.160.001	6.200.000	6.200.001	6.280.000
	6.280.001	6.290.000	6.290.001	6.330.000
	6.330.001	6.350.000	6.350.001	6.370.000
	6.370.001	6.380.000	6.380.001	6.400.000
	6.400.001	6.450.000	6.450.001	6.470.000
	6.470.001	6.490.000	6.490.001	6.510.000
	6.510.001	6.530.000	6.530.001	6.560.000
	6.560.001	6.570.000	6.570.001	6.630.000
	6.630.001	6.690.000	6.690.001	6.710.000
	6.710.001	6.730.000	6.730.001	6.770.000
	6.770.001	6.800.000	6.800.001	6.850.000
	6.850.001	6.860.000	6.860.001	6.870.000
	6.870.001	6.890.000	6.890.001	6.900.000
	6.900.001	6.970.000	6.970.001	6.990.000
	6.990.001	7.000.000	7.000.001	7.030.000
	7.030.001	7.040.000	7.040.001	7.180.000
7.180.001	7.190.000	7.190.001	7.220.000	
7.220.001	7.230.000	7.230.001	7.240.000	
7.240.001	7.260.000	7.260.001	7.270.000	
7.270.001	7.400.000	7.400.001	7.410.000	
7.410.001	7.430.000	7.430.001	7.450.000	

<i>10kb</i>	$\rho$			
	LSB		NLSB	
	Inicio	Fin	Inicio	Fin
<i>2L</i>	7.450.001	7.480.000	7.480.001	7.490.000
	7.490.001	7.630.000	7.630.001	7.640.000
	7.640.001	7.680.000	7.680.001	7.690.000
	7.690.001	7.830.000	7.830.001	7.840.000
	7.840.001	7.960.000	7.960.001	7.970.000
	7.970.001	8.000.000	8.000.001	8.010.000
	8.010.001	8.020.000	8.020.001	8.050.000
	8.050.001	8.140.000	8.140.001	8.150.000
	8.150.001	8.160.000	8.160.001	8.190.000
	8.190.001	8.220.000	8.220.001	8.240.000
	8.240.001	8.250.000	8.250.001	8.260.000
	8.260.001	8.320.000	8.320.001	8.350.000
	8.350.001	8.490.000	8.490.001	8.510.000
	8.510.001	8.530.000	8.530.001	8.560.000
	8.560.001	8.580.000	8.580.001	8.590.000
	8.590.001	8.610.000	8.610.001	8.670.000
	8.670.001	8.740.000	8.740.001	8.750.000
	8.750.001	8.960.000	8.960.001	8.970.000
	8.970.001	8.980.000	8.980.001	8.990.000
	8.990.001	9.110.000	9.110.001	9.120.000
	9.120.001	9.130.000	9.130.001	9.220.000
	9.220.001	9.350.000	9.350.001	9.370.000
	9.370.001	9.410.000	9.410.001	9.420.000
	9.420.001	9.440.000	9.440.001	9.520.000
	9.520.001	9.530.000	9.530.001	9.540.000
	9.540.001	9.550.000	9.550.001	9.560.000
	9.560.001	9.640.000	9.640.001	9.710.000
	9.710.001	9.720.000	9.720.001	9.730.000
	9.730.001	9.770.000	9.770.001	9.780.000
	9.780.001	9.810.000	9.810.001	9.890.000
9.890.001	9.920.000	9.920.001	9.950.000	
9.950.001	9.980.000	9.980.001	10.000.000	
10.000.001	10.010.000	10.010.001	10.220.000	
10.220.001	10.230.000	10.230.001	10.260.000	
10.260.001	10.280.000	10.280.001	10.330.000	

<i>10kb</i>	$\rho$			
	LSB		NLSB	
	Inicio	Fin	Inicio	Fin
<i>2L</i>	10.330.001	10.340.000	10.340.001	10.380.000
	10.380.001	10.410.000	10.410.001	10.430.000
	10.430.001	10.440.000	10.440.001	10.590.000
	10.590.001	10.600.000	10.600.001	10.640.000
	10.640.001	10.650.000	10.650.001	10.680.000
	10.680.001	10.750.000	10.750.001	10.760.000
	10.760.001	10.780.000	10.780.001	10.800.000
	10.800.001	10.810.000	10.810.001	10.840.000
	10.840.001	10.880.000	10.880.001	10.940.000
	10.940.001	11.070.000	11.070.001	11.080.000
	11.080.001	11.140.000	11.140.001	11.150.000
	11.150.001	11.170.000	11.170.001	11.210.000
	11.210.001	11.300.000	11.300.001	11.360.000
	11.360.001	11.370.000	11.370.001	11.380.000
	11.380.001	11.420.000	11.420.001	11.430.000
	11.430.001	11.550.000	11.550.001	11.560.000
	11.560.001	11.780.000	11.780.001	11.790.000
	11.790.001	11.820.000	11.820.001	11.830.000
	11.830.001	11.880.000	11.880.001	11.890.000
	11.890.001	11.930.000	11.930.001	11.940.000
	11.940.001	11.960.000	11.960.001	11.980.000
	11.980.001	12.020.000	12.020.001	12.030.000
	12.030.001	12.080.000	12.080.001	12.090.000
	12.090.001	12.120.000	12.120.001	12.130.000
	12.130.001	12.140.000	12.140.001	12.210.000
	12.210.001	12.230.000	12.230.001	12.270.000
	12.270.001	12.280.000	12.280.001	12.350.000
	12.350.001	12.360.000	12.360.001	12.420.000
	12.420.001	12.460.000	12.460.001	12.480.000
	12.480.001	12.490.000	12.490.001	12.520.000
12.520.001	12.530.000	12.530.001	12.560.000	
12.560.001	12.570.000	12.530.001	12.580.000	
12.580.001	12.630.000	12.630.001	12.650.000	
12.650.001	13.050.000	13.050.001	13.060.000	
13.060.001	13.100.000	13.100.001	13.140.000	

<i>10kb</i>	$\rho$			
	LSB		NLSB	
	Inicio	Fin	Inicio	Fin
<i>2L</i>	13.140.001	13.310.000	13.310.001	13.350.000
	13.350.001	13.530.000	13.530.001	13.540.000
	13.540.001	13.610.000	13.610.001	13.620.000
	13.620.001	13.670.000	13.670.001	13.770.000
	13.770.001	13.800.000	13.800.001	13.810.000
	13.810.001	13.880.000	13.880.001	13.890.000
	13.890.001	13.920.000	13.920.001	13.930.000
	13.930.001	13.940.000	13.940.001	13.950.000
	13.950.001	13.960.000	13.960.001	13.970.000
	13.970.001	13.980.000	13.980.001	14.040.000
	14.040.001	14.050.000	14.050.001	14.090.000
	14.090.001	14.220.000	14.220.001	14.360.000
	14.360.001	14.370.000	14.370.001	14.400.000
	14.400.001	14.430.000	14.430.001	14.460.000
	14.460.001	14.960.000	14.960.001	14.980.000
	14.980.001	15.000.000	15.000.001	15.040.000
	15.040.001	15.080.000	15.080.001	15.110.000
	15.110.001	15.120.000	15.120.001	15.130.000
	15.130.001	15.160.000	15.160.001	15.180.000
	15.180.001	15.190.000	15.190.001	15.200.000
	15.200.001	15.220.000	15.220.001	15.230.000
	15.230.001	15.240.000	15.240.001	15.260.000
	15.260.001	15.270.000	15.270.001	15.340.000
	15.340.001	15.350.000	15.350.001	15.360.000
	15.360.001	15.370.000	15.370.001	15.410.000
	15.410.001	15.430.000	15.430.001	15.440.000
	15.440.001	15.460.000	15.460.001	15.480.000
	15.480.001	15.730.000	15.730.001	15.760.000
	15.760.001	15.770.000	15.770.001	15.780.000
	15.780.001	16.100.000	16.100.001	16.130.000
	16.130.001	16.180.000	16.180.001	16.240.000
	16.240.001	16.290.000	16.290.001	16.300.000
16.300.001	16.360.000	16.360.001	16.420.000	
16.420.001	16.430.000	16.430.001	16.440.000	
16.440.001	16.450.000	16.450.001	16.480.000	

<i>10kb</i>	$\rho$			
	LSB		NLSB	
	Inicio	Fin	Inicio	Fin
<i>2L</i>	16.480.001	16.490.000	16.490.001	16.510.000
	16.510.001	16.550.000	16.550.001	16.580.000
	16.580.001	16.590.000	16.590.001	16.620.000
	16.620.001	17.390.000	17.390.001	17.410.000
	17.410.001	17.490.000	17.490.001	17.500.000
	17.500.001	18.570.000	18.570.001	18.580.000
	18.580.001	18.640.000	18.640.001	18.650.000
	18.650.001	18.800.000	18.800.001	18.810.000
	18.810.001	18.820.000	18.820.001	18.830.000
	18.830.001	18.860.000	18.860.001	18.890.000
	18.890.001	18.900.000	18.900.001	18.920.000
	18.920.001	19.590.000	19.590.001	19.600.000
	19.600.001	19.610.000	19.610.001	19.620.000
	19.620.001	19.810.000	19.810.001	19.820.000
	19.820.001	20.440.000	20.440.001	20.460.000
	20.460.001	21.240.000	21.240.001	21.250.000
	21.250.001	21.630.000	21.630.001	21.640.000
	21.640.001	22.500.000		

<i>10kb</i>	$\rho$			
	LSB		NLSB	
	Inicio	Fin	Inicio	Fin
<i>2R</i>	1	340.000	340.001	350.000
	350.001	1.570.000	1.570.001	1.580.000
	1.580.001	2.240.000	2.240.001	2.250.000
	2.250.001	2.320.000	2.320.001	2.330.000
	2.330.001	3.730.000	3.730.001	3.740.000
	3.740.001	4.480.000	4.480.001	4.500.000
	4.500.001	4.520.000	4.520.001	4.530.000
	4.530.001	5.300.000	5.300.001	5.310.000
	5.310.001	5.340.000	5.340.001	5.350.000
	5.350.001	6.230.000	6.230.001	6.240.000
	6.240.001	6.330.000	6.330.001	6.350.000
	6.350.001	6.440.000	6.440.001	6.450.000
	6.450.001	6.540.000	6.540.001	6.550.000
	6.550.001	6.560.000	6.560.001	6.590.000
	6.590.001	6.640.000	6.640.001	6.650.000
	6.650.001	6.660.000	6.660.001	6.670.000
	6.670.001	6.730.000	6.730.001	6.760.000
	6.760.001	6.790.000	6.790.001	6.830.000
	6.830.001	6.980.000	6.980.001	6.990.000
	6.990.001	7.030.000	7.030.001	7.040.000
	7.040.001	7.050.000	7.050.001	7.070.000
	7.070.001	7.260.000	7.260.001	7.270.000
	7.270.001	7.290.000	7.290.001	7.300.000
	7.300.001	7.310.000	7.310.001	7.330.000
	7.330.001	7.440.000	7.440.001	7.460.000
	7.460.001	7.960.000	7.960.001	7.970.000
	7.970.001	8.140.000	8.140.001	8.150.000
	8.150.001	8.330.000	8.330.001	8.340.000
	8.340.001	8.420.000	8.420.001	8.430.000
	8.430.001	8.490.000	8.490.001	8.500.000
8.500.001	8.580.000	8.580.001	8.590.000	
8.590.001	8.610.000	8.610.001	8.620.000	
8.620.001	8.890.000	8.890.001	8.910.000	
8.910.001	8.930.000	8.930.001	8.940.000	
8.940.001	8.960.000	8.960.001	8.970.000	



<i>10kb</i>	$\rho$			
	LSB		NLSB	
	Inicio	Fin	Inicio	Fin
<i>2R</i>	8.970.001	9.160.000	9.160.001	9.170.000
	9.170.001	9.260.000	9.260.001	9.280.000
	9.280.001	9.320.000	9.320.001	9.330.000
	9.330.001	9.340.000	9.340.001	9.370.000
	9.370.001	9.410.000	9.410.001	9.420.000
	9.420.001	9.520.000	9.520.001	9.530.000
	9.530.001	9.560.000	9.560.001	9.570.000
	9.570.001	9.600.000	9.600.001	9.610.000
	9.610.001	9.630.000	9.630.001	9.640.000
	9.640.001	9.660.000	9.660.001	9.670.000
	9.670.001	9.930.000	9.930.001	9.940.000
	9.940.001	9.970.000	9.970.001	9.980.000
	9.980.001	10.000.000	10.000.001	10.010.000
	10.010.001	10.510.000	10.510.001	10.530.000
	10.530.001	10.570.000	10.570.001	10.640.000
	10.640.001	10.660.000	10.660.001	10.680.000
	10.680.001	10.690.000	10.690.001	10.720.000
	10.720.001	10.890.000	10.890.001	10.900.000
	10.900.001	10.930.000	10.930.001	10.940.000
	10.940.001	10.980.000	10.980.001	10.990.000
	10.990.001	11.720.000	11.720.001	11.740.000
	11.740.001	11.760.000	11.760.001	11.810.000
	11.810.001	11.940.000	11.940.001	11.950.000
	11.950.001	11.960.000	11.960.001	11.990.000
	11.990.001	12.000.000	12.000.001	12.010.000
	12.010.001	12.040.000	12.040.001	12.070.000
	12.070.001	12.090.000	12.090.001	12.100.000
	12.100.001	12.110.000	12.110.001	12.150.000
	12.150.001	12.190.000	12.190.001	12.200.000
	12.200.001	12.240.000	12.240.001	12.250.000
12.250.001	12.310.000	12.310.001	12.380.000	
12.380.001	12.390.000	12.390.001	12.400.000	
12.400.001	12.440.000	12.440.001	12.460.000	
12.460.001	12.470.000	12.470.001	12.490.000	
12.490.001	12.690.000	12.690.001	12.700.000	

<i>10kb</i>	$\rho$			
	LSB		NLSB	
	Inicio	Fin	Inicio	Fin
<i>2R</i>	12.700.001	12.860.000	12.860.001	12.870.000
	12.870.001	13.360.000	13.360.001	13.370.000
	13.370.001	13.500.000	13.500.001	13.510.000
	13.510.001	13.530.000	13.530.001	13.550.000
	13.550.001	13.590.000	13.590.001	13.600.000
	13.600.001	13.690.000	13.690.001	13.730.000
	13.730.001	13.740.000	13.740.001	13.750.000
	13.750.001	13.760.000	13.760.001	13.790.000
	13.790.001	13.800.000	13.800.001	13.840.000
	13.840.001	13.850.000	13.850.001	13.870.000
	13.870.001	13.910.000	13.910.001	13.930.000
	13.930.001	13.940.000	13.940.001	13.950.000
	13.950.001	14.080.000	14.080.001	14.090.000
	14.090.001	14.140.000	14.140.001	14.150.000
	14.150.001	14.260.000	14.260.001	14.280.000
	14.280.001	14.450.000	14.450.001	14.460.000
	14.460.001	14.530.000	14.530.001	14.540.000
	14.540.001	14.550.000	14.550.001	14.570.000
	14.570.001	14.580.000	14.580.001	14.590.000
	14.590.001	14.610.000	14.610.001	14.620.000
	14.620.001	14.630.000	14.630.001	14.640.000
	14.640.001	14.700.000	14.700.001	14.720.000
	14.720.001	14.780.000	14.780.001	14.790.000
	14.790.001	14.810.000	14.810.001	14.820.000
	14.820.001	14.830.000	14.830.001	14.850.000
	14.850.001	14.980.000	14.980.001	15.020.000
	15.020.001	15.040.000	15.040.001	15.110.000
	15.110.001	15.130.000	15.130.001	15.140.000
	15.140.001	15.220.000	15.220.001	15.230.000
	15.230.001	15.260.000	15.260.001	15.280.000
15.280.001	15.380.000	15.380.001	15.410.000	
15.410.001	16.020.000	16.020.001	16.040.000	
16.040.001	16.050.000	16.050.001	16.100.000	
16.100.001	16.180.000	16.180.001	16.190.000	
16.190.001	16.230.000	16.230.001	16.240.000	

<i>10kb</i>	$\rho$			
	LSB		NLSB	
	Inicio	Fin	Inicio	Fin
<i>2R</i>	16.230.001	16.240.000	16.250.001	16.260.000
	16.260.001	16.270.000	16.270.001	16.280.000
	16.280.001	16.290.000	16.290.001	16.330.000
	16.330.001	16.340.000	16.340.001	16.390.000
	16.390.001	16.400.000	16.400.001	16.420.000
	16.420.001	16.430.000	16.430.001	16.470.000
	16.470.001	16.480.000	16.480.001	16.500.000
	16.500.001	16.560.000	16.560.001	16.570.000
	16.570.001	16.620.000	16.620.001	16.630.000
	16.630.001	16.810.000	16.810.001	16.860.000
	16.860.001	16.870.000	16.870.001	16.880.000
	16.880.001	16.910.000	16.910.001	16.940.000
	16.940.001	16.950.000	16.950.001	16.960.000
	16.960.001	16.980.000	16.980.001	17.010.000
	17.010.001	17.020.000	17.020.001	17.030.000
	17.030.001	17.060.000	17.060.001	17.070.000
	17.070.001	17.080.000	17.080.001	17.100.000
	17.100.001	17.110.000	17.110.001	17.170.000
	17.170.001	17.290.000	17.290.001	17.350.000
	17.350.001	17.370.000	17.370.001	17.380.000
	17.380.001	17.390.000	17.390.001	17.440.000
	17.440.001	17.460.000	17.460.001	17.500.000
	17.500.001	17.560.000	17.560.001	17.570.000
	17.570.001	17.580.000	17.580.001	17.590.000
	17.590.001	17.610.000	17.610.001	17.620.000
	17.620.001	17.640.000	17.640.001	17.650.000
	17.650.001	17.660.000	17.660.001	17.690.000
	17.690.001	17.730.000	17.730.001	17.750.000
	17.750.001	17.760.000	17.760.001	17.770.000
	17.770.001	17.780.000	17.780.001	17.810.000
17.810.001	17.840.000	17.840.001	17.860.000	
17.860.001	17.940.000	17.940.001	17.960.000	
17.960.001	18.030.000	18.030.001	18.040.000	
18.040.001	18.270.000	18.270.001	18.290.000	
18.290.001	18.320.000	18.320.001	18.350.000	

<i>10kb</i>	$\rho$			
	LSB		NLSB	
	Inicio	Fin	Inicio	Fin
<i>2R</i>	18.350.001	18390.000	18.390.001	18.400.000
	18.400.001	18.420.000	18.420.001	18.480.000
	18.480.001	18.580.000	18.580.001	18.600.000
	18.600.001	18.890.000	18.890.001	18.900.000
	18.900.001	18.920.000	18.920.001	18.930.000
	18.930.001	18.970.000	18.970.001	18.990.000
	18.990.001	19.000.000	19.000.001	19.030.000
	19.030.001	19.040.000	19.040.001	19.060.000
	19.060.001	19.070.000	19.070.001	19.080.000
	19.080.001	19.110.000	19.110.001	19.140.000
	19.140.001	19.190.000	19.190.001	19.200.000
	19.200.001	19.210.000	19.210.001	19.220.000
	19.220.001	19.300.000	19.300.001	19.310.000
	19.310.001	19.320.000	19.320.001	19.350.000
	19.350.001	19.390.000	19.390.001	19.400.000
	19.400.001	19.450.000	19.450.001	19.460.000
	19.460.001	19.480.000	19.480.001	19.560.000
	19.560.001	19.610.000	19.610.001	19.620.000
	19.620.001	19.680.000	19.680.001	19.690.000
	19.690.001	19.950.000	19.950.001	19.970.000
19.970.001	21.140.000			

<i>10kb</i>	$\rho$			
	LSB		NLSB	
	Inicio	Fin	Inicio	Fin
<i>3L</i>	1	110.000	110.001	120.000
	120.001	600.000	600.001	610.000
	610.001	660.000	660.001	680.000
	680.001	820.000	820.001	830.000
	830.001	840.000	840.001	860.000
	860.001	900.000	900.001	920.000
	920.001	940.000	940.001	950.000
	950.001	1.000.000	1.000.001	1.020.000
	1.020.001	1.040.000	1.040.001	1.050.000
	1.050.001	1.060.000	1.060.001	1.090.000
	1.090.001	1.110.000	1.110.001	1.120.000
	1.120.001	1.350.000	1.350.001	1.420.000
	1.420.001	1.430.000	1.430.001	1.470.000
	1.470.001	1.520.000	1.520.001	1.530.000
	1.530.001	1.580.000	1.580.001	1.590.000
	1.590.001	1.610.000	1.610.001	1.620.000
	1.620.001	1.670.000	1.670.001	1.700.000
	1.700.001	1.920.000	1.920.001	1.930.000
	1.930.001	2.120.000	2.120.001	2.130.000
	2.130.001	2.140.000	2.140.001	2.160.000
	2.160.001	2.290.000	2.290.001	2.300.000
	2.300.001	2.310.000	2.310.001	2.370.000
	2.370.001	2.380.000	2.380.001	2.410.000
	2.410.001	2.420.000	2.420.001	2.440.000
	2.440.001	2.510.000	2.510.001	2.580.000
	2.580.001	2.600.000	2.600.001	2.630.000
	2.630.001	2.680.000	2.680.001	2.700.000
	2.700.001	2.830.000	2.830.001	2.850.000
	2.850.001	2.860.000	2.860.001	2.890.000
	2.890.001	2.990.000	2.990.001	3.010.000
	3.010.001	3.020.000	3.020.001	3.030.000
	3.030.001	3.100.000	3.100.001	3.120.000
3.120.001	3.130.000	3.130.001	3.140.000	
3.140.001	3.260.000	3.260.001	3.270.000	
3.270.001	3.290.000	3.290.001	3.300.000	

<i>10kb</i>	$\rho$			
	LSB		NLSB	
	Inicio	Fin	Inicio	Fin
<i>3L</i>	3.300.001	3.450.000	3.450.001	3.460.000
	3.460.001	3.480.000	3.480.001	3.510.000
	3.510.001	3.520.000	3.520.001	3.540.000
	3.540.001	3.560.000	3.560.001	3.580.000
	3.580.001	3.590.000	3.590.001	3.600.000
	3.600.001	3.620.000	3.620.001	3.630.000
	3.630.001	3.710.000	3.710.001	3.750.000
	3.750.001	3.770.000	3.770.001	3.810.000
	3.810.001	3.850.000	3.850.001	3.870.000
	3.870.001	3.880.000	3.880.001	3.890.000
	3.890.001	3.910.000	3.910.001	3.920.000
	3.920.001	3.950.000	3.950.001	3.960.000
	3.960.001	3.980.000	3.980.001	4.000.000
	4.000.001	4.010.000	4.010.001	4.030.000
	4.030.001	4.040.000	4.040.001	4.100.000
	4.100.001	4.110.000	4.110.001	4.120.000
	4.120.001	4.340.000	4.340.001	4.350.000
	4.350.001	4.380.000	4.380.001	4.400.000
	4.400.001	4.520.000	4.520.001	4.540.000
	4.540.001	4.560.000	4.560.001	4.570.000
	4.570.001	4.600.000	4.600.001	4.620.000
	4.620.001	4.640.000	4.640.001	4.670.000
	4.670.001	4.680.000	4.680.001	4.690.000
	4.690.001	4.700.000	4.700.001	4.740.000
	4.740.001	4.750.000	4.750.001	4.770.000
	4.770.001	4.910.000	4.910.001	4.920.000
	4.920.001	5.000.000	5.000.001	5.010.000
	5.010.001	5.210.000	5.210.001	5.220.000
	5.220.001	5.260.000	5.260.001	5.270.000
	5.270.001	5.280.000	5.280.001	5.290.000
5.290.001	5.300.000	5.300.001	5.320.000	
5.320.001	5.370.000	5.370.001	5.400.000	
5.400.001	5.410.000	5.410.001	5.440.000	
5.440.001	5.450.000	5.450.001	5.470.000	
5.470.001	5.480.000	5.480.001	5.490.000	

<i>10kb</i>	$\rho$			
	LSB		NLSB	
	Inicio	Fin	Inicio	Fin
<i>3L</i>	5.490.001	5.510.000	5.510.001	5.530.000
	5.530.001	5.540.000	5.540.001	5.550.000
	5.550.001	5.790.000	5.790.001	5.800.000
	5.800.001	5.950.000	5.950.001	5.960.000
	5.960.001	5.980.000	5.980.001	5.990.000
	5.990.001	6.460.000	6.460.001	6.470.000
	6.470.001	6.530.000	6.530.001	6.540.000
	6.540.001	6.570.000	6.570.001	6.580.000
	6.580.001	6.600.000	6.600.001	6.610.000
	6.610.001	6.970.000	6.970.001	7.000.000
	7.000.001	7.010.000	7.010.001	7.020.000
	7.020.001	7.090.000	7.090.001	7.110.000
	7.110.001	7.200.000	7.200.001	7.210.000
	7.210.001	7.220.000	7.220.001	7.230.000
	7.230.001	7.250.000	7.250.001	7.300.000
	7.300.001	7.410.000	7.410.001	7.430.000
	7.430.001	7.460.000	7.460.001	7.480.000
	7.480.001	7.580.000	7.580.001	7.590.000
	7.590.001	7.650.000	7.650.001	7.660.000
	7.660.001	7.870.000	7.870.001	7.880.000
	7.880.001	7.890.000	7.890.001	7.900.000
	7.900.001	7.930.000	7.930.001	7.940.000
	7.940.001	7.950.000	7.950.001	7.960.000
	7.960.001	7.980.000	7.980.001	8.010.000
	8.010.001	8.020.000	8.020.001	8.050.000
	8.050.001	8.060.000	8.060.001	8.090.000
	8.090.001	8.150.000	8.150.001	8.160.000
	8.160.001	8.210.000	8.210.001	8.220.000
	8.220.001	8.230.000	8.230.001	8.270.000
	8.270.001	8.310.000	8.310.001	8.330.000
	8.330.001	8.410.000	8.410.001	8.420.000
	8.420.001	8.430.000	8.430.001	8.440.000
8.440.001	8.450.000	8.450.001	8.460.000	
8.460.001	8.540.000	8.540.001	8.550.000	
8.550.001	8.570.000	8.570.001	8.580.000	

<i>10kb</i>	$\rho$			
	LSB		NLSB	
	Inicio	Fin	Inicio	Fin
<i>3L</i>	8.580.001	8.600.000	8.600.001	8.610.000
	8.610.001	8.680.000	8.680.001	8.690.000
	8.690.001	8.700.000	8.700.001	8.710.000
	8.710.001	8.850.000	8.850.001	8.870.000
	8.870.001	8.890.000	8.890.001	8.900.000
	8.900.001	8.910.000	8.910.001	8.920.000
	8.920.001	9.020.000	9.020.001	9.030.000
	9.030.001	9.080.000	9.080.001	9.090.000
	9.090.001	9.100.000	9.100.001	9.110.000
	9.110.001	9.140.000	9.140.001	9.150.000
	9.150.001	9.160.000	9.160.001	9.170.000
	9.170.001	9.190.000	9.190.001	9.260.000
	9.260.001	9.270.000	9.270.001	9.290.000
	9.290.001	9.300.000	9.300.001	9.320.000
	9.320.001	9.350.000	9.350.001	9.360.000
	9.360.001	9.490.000	9.490.001	9.500.000
	9.500.001	9.510.000	9.510.001	9.520.000
	9.520.001	9.540.000	9.540.001	9.570.000
	9.570.001	9.760.000	9.760.001	9.770.000
	9.770.001	9.800.000	9.800.001	9.810.000
	9.810.001	10.260.000	10.260.001	10.270.000
	10.270.001	10.330.000	10.330.001	10.350.000
	10.350.001	10.360.000	10.360.001	10.380.000
	10.380.001	10.410.000	10.410.001	10.470.000
	10.470.001	10.500.000	10.500.001	10.510.000
	10.510.001	10.690.000	10.690.001	10.830.000
	10.830.001	10.880.000	10.880.001	10.940.000
	10.940.001	10.980.000	10.980.001	10.990.000
	10.990.001	11.000.000	11.000.001	11.010.000
	11.010.001	11.020.000	11.020.001	11.040.000
	11.040.001	11.140.000	11.140.001	11.160.000
	11.160.001	11.170.000	11.170.001	11.180.000
11.180.001	11.230.000	11.230.001	11.260.000	
11.260.001	11.300.000	11.300.001	11.420.000	
11.420.001	11.430.000	11.430.001	11.480.000	



<i>10kb</i>	$\rho$			
	LSB		NLSB	
	Inicio	Fin	Inicio	Fin
<i>3L</i>	11.480.001	11.500.000	11.500.001	11.540.000
	11.540.001	11.790.000	11.790.001	11.800.000
	11.800.001	11.860.000	11.860.001	11.870.000
	11.870.001	11.880.000	11.880.001	11.910.000
	11.910.001	12.010.000	12.010.001	12.070.000
	12.070.001	12.080.000	12.080.001	12.100.000
	12.100.001	12.310.000	12.310.001	12.320.000
	12.320.001	12.350.000	12.350.001	12.360.000
	12.360.001	12.370.000	12.370.001	12.380.000
	12.380.001	12.400.000	12.400.001	12.420.000
	12.420.001	12.430.000	12.430.001	12.460.000
	12.460.001	12.470.000	12.470.001	12.730.000
	12.730.001	12.770.000	12.770.001	12.780.000
	12.780.001	12.790.000	12.790.001	12.800.000
	12.800.001	12.820.000	12.820.001	12.830.000
	12.830.001	12.840.000	12.840.001	12.860.000
	12.860.001	12.870.000	12.870.001	12.900.000
	12.900.001	12.920.000	12.920.001	12.940.000
	12.940.001	12.970.000	12.970.001	12.990.000
	12.990.001	13.030.000	13.030.001	13.040.000
	13.040.001	13.110.000	13.110.001	13.130.000
	13.130.001	13.140.000	13.140.001	13.180.000
	13.180.001	13.210.000	13.210.001	13.230.000
	13.230.001	13.250.000	13.250.001	13.260.000
	13.260.001	13.270.000	13.270.001	13.280.000
	13.280.001	13.900.000	13.900.001	13.910.000
	13.910.001	13.930.000	13.930.001	13.950.000
	13.950.001	13.970.000	13.970.001	13.980.000
	13.980.001	14.080.000	14.080.001	14.100.000
	14.100.001	14.140.000	14.140.001	14.160.000
14.160.001	14.200.000	14.200.001	14.210.000	
14.210.001	14.300.000	14.300.001	14.310.000	
14.310.001	14.760.000	14.760.001	14.770.000	
14.770.001	14.800.000	14.800.001	14.810.000	
14.810.001	15.000.000	15.000.001	15.020.000	

<i>10kb</i>	$\rho$			
	LSB		NLSB	
	Inicio	Fin	Inicio	Fin
<i>3L</i>	15.020.001	15.050.000	15.050.001	15.060.000
	15.060.001	15.110.000	15.110.001	15.120.000
	15.120.001	15.170.000	15.170.001	15.180.000
	15.180.001	15.360.000	15.360.001	15.380.000
	15.380.001	15.590.000	15.590.001	15.600.000
	15.600.001	15.750.000	15.750.001	15.760.000
	15.760.001	16.020.000	16.020.001	16.030.000
	16.030.001	16.050.000	16.050.001	16.060.000
	16.060.001	16.410.000	16.410.001	16.430.000
	16.430.001	16.530.000	16.530.001	16.560.000
	16.560.001	16.590.000	16.590.001	16.600.000
	16.600.001	16.620.000	16.620.001	16.630.000
	16.630.001	16.690.000	16.690.001	16.700.000
	16.700.001	16.740.000	16.740.001	16.760.000
	16.760.001	16.860.000	16.860.001	16.890.000
	16.890.001	16.900.000	16.900.001	16.910.000
	16.910.001	17.280.000	17.280.001	17.290.000
	17.290.001	18.040.000	18.040.001	18.050.000
	18.050.001	18.090.000	18.090.001	18.100.000
	18.100.001	18.110.000	18.110.001	18.120.000
	18.120.001	18.170.000	18.170.001	18.180.000
	18.180.001	18.190.000	18.190.001	18.220.000
	18.220.001	18.270.000	18.270.001	18.280.000
	18.280.001	18.300.000	18.300.001	18.310.000
	18.310.001	18.320.000	18.320.001	18.330.000
	18.330.001	18.600.000	18.600.001	18.610.000
	18.610.001	18.840.000	18.840.001	18.850.000
	18.850.001	19.010.000	19.010.001	19.020.000
	19.020.001	19.660.000	19.660.001	19.670.000
	19.670.001	19.930.000	19.930.001	19.940.000
19.940.001	24.130.000	24.130.001	24.140.000	
24.140.001	24.540.000			

<i>10kb</i>	$\rho$			
	LSB		NLSB	
	Inicio	Fin	Inicio	Fin
<i>3R</i>	1	650.000	650.001	660.000
	660.001	4.020.000	4.020.001	4.030.000
	4.030.001	4.100.000	4.100.001	4.120.000
	4.120.001	4.130.000	4.130.001	4.140.000
	4.140.001	4.280.000	4.280.001	4.300.000
	4.300.001	4.370.000	4.370.001	4.380.000
	4.380.001	4.440.000	4.440.001	4.450.000
	4.450.001	4.710.000	4.710.001	4.720.000
	4.720.001	5.120.000	5.120.001	5.150.000
	5.150.001	5.190.000	5.190.001	5.210.000
	5.210.001	5.310.000	5.310.001	5.320.000
	5.320.001	5.510.000	5.510.001	5.520.000
	5.520.001	5.690.000	5.690.001	5.700.000
	5.700.001	5.900.000	5.900.001	5.910.000
	5.910.001	6.490.000	6.490.001	6.510.000
	6.510.001	7.870.000	7.870.001	7.880.000
	7.880.001	8.160.000	8.160.001	8.170.000
	8.170.001	8.210.000	8.210.001	8.220.000
	8.220.001	8.390.000	8.390.001	8.420.000
	8.420.001	8.430.000	8.430.001	8.440.000
	8.440.001	8.550.000	8.550.001	8.560.000
	8.560.001	9.320.000	9.320.001	9.330.000
	9.330.001	9.370.000	9.370.001	9.380.000
	9.380.001	9.490.000	9.490.001	9.500.000
	9.500.001	9.530.000	9.530.001	9.560.000
	9.560.001	9.570.000	9.570.001	9.590.000
	9.590.001	9.610.000	9.610.001	9.620.000
	9.620.001	9.640.000	9.640.001	9.660.000
	9.660.001	9.680.000	9.680.001	9.690.000
	9.690.001	9.710.000	9.710.001	9.720.000
	9.720.001	9.740.000	9.740.001	9.800.000
	9.800.001	9.810.000	9.810.001	9.840.000
9.840.001	9.860.000	9.860.001	9.920.000	
9.920.001	9.960.000	9.960.001	10.000.000	
10.000.001	10.030.000	10.030.001	10.060.000	

<i>10kb</i>	$\rho$			
	LSB		NLSB	
	Inicio	Fin	Inicio	Fin
<i>3R</i>	10.060.001	10.070.000	10.070.001	10.080.000
	10.080.001	10.110.000	10.110.001	10.150.000
	10.150.001	10.160.000	10.160.001	10.230.000
	10.230.001	10.250.000	10.250.001	10.300.000
	10.300.001	10.410.000	10.410.001	10.430.000
	10.430.001	10.450.000	10.450.001	10.460.000
	10.460.001	10.490.000	10.490.001	10.520.000
	10.520.001	10.530.000	10.530.001	10.540.000
	10.540.001	10.550.000	10.550.001	10.560.000
	10.560.001	10.570.000	10.570.001	10.620.000
	10.620.001	10.730.000	10.730.001	10.740.000
	10.740.001	10.770.000	10.770.001	10.780.000
	10.780.001	10.820.000	10.820.001	10.830.000
	10.830.001	10.840.000	10.840.001	10.850.000
	10.850.001	10.870.000	10.870.001	10.920.000
	10.920.001	10.940.000	10.940.001	10.950.000
	10.950.001	10.990.000	10.990.001	11.000.000
	11.000.001	11.120.000	11.120.001	11.140.000
	11.140.001	11.210.000	11.210.001	11.230.000
	11.230.001	11.250.000	11.250.001	11.310.000
	11.310.001	11.340.000	11.340.001	11.350.000
	11.350.001	11.480.000	11.480.001	11.490.000
	11.490.001	11.570.000	11.570.001	11.590.000
	11.590.001	11.650.000	11.650.001	11.660.000
	11.660.001	11.690.000	11.690.001	11.710.000
	11.710.001	12.000.000	12.000.001	12.010.000
	12.010.001	12.020.000	12.020.001	12.050.000
	12.050.001	12.930.000	12.930.001	12.960.000
	12.960.001	13.280.000	13.280.001	13.290.000
	13.290.001	13.310.000	13.310.001	13.320.000
	13.320.001	13.410.000	13.410.001	13.420.000
	13.420.001	13.450.000	13.450.001	13.470.000
13.470.001	13.480.000	13.480.001	13.490.000	
13.490.001	13.550.000	13.550.001	13.590.000	
13.590.001	13.640.000	13.640.001	13.660.000	

<i>10kb</i>	$\rho$			
	LSB		NLSB	
	Inicio	Fin	Inicio	Fin
<i>3R</i>	13.660.001	13.680.000	13.680.001	13.700.000
	13.700.001	13.730.000	13.730.001	13.740.000
	13.740.001	13.750.000	13.750.001	13.990.000
	13.990.001	14.000.000	14.000.001	14.010.000
	14.010.001	14.020.000	14.020.001	14.050.000
	14.050.001	14.060.000	14.060.001	14.070.000
	14.070.001	14.100.000	14.100.001	14.120.000
	14.120.001	14.130.000	14.130.001	14.180.000
	14.180.001	14.240.000	14.240.001	14.350.000
	14.350.001	14.360.000	14.360.001	14.400.000
	14.400.001	14.420.000	14.420.001	14.470.000
	14.470.001	14.630.000	14.630.001	14.740.000
	14.740.001	14.760.000	14.760.001	14.800.000
	14.800.001	14.810.000	14.810.001	14.840.000
	14.840.001	14.870.000	14.870.001	14.910.000
	14.910.001	14.920.000	14.920.001	14.980.000
	14.980.001	14.990.000	14.990.001	15.020.000
	15.020.001	15.030.000	15.030.001	15.040.000
	15.040.001	15.070.000	15.070.001	15.140.000
	15.140.001	15.160.000	15.160.001	15.170.000
	15.170.001	15.210.000	15.210.001	15.220.000
	15.220.001	15.240.000	15.240.001	15.250.000
	15.250.001	15.260.000	15.260.001	15.270.000
	15.270.001	15.280.000	15.280.001	15.290.000
	15.290.001	15.300.000	15.300.001	15.310.000
	15.310.001	15.580.000	15.580.001	15.590.000
	15.590.001	15.640.000	15.640.001	15.660.000
	15.660.001	16.240.000	16.240.001	16.250.000
	16.250.001	16.270.000	16.270.001	16.280.000
	16.280.001	16.390.000	16.390.001	16.400.000
16.400.001	16.480.000	16.480.001	16.490.000	
16.490.001	16.510.000	16.510.001	16.520.000	
16.520.001	16.750.000	16.750.001	16.770.000	
16.770.001	16.890.000	16.890.001	16.900.000	
16.900.001	16.980.000	16.980.001	17.010.000	

<i>10kb</i>	$\rho$			
	LSB		NLSB	
	Inicio	Fin	Inicio	Fin
<i>3R</i>	17.010.001	17.060.000	17.060.001	17.080.000
	17.080.001	17.330.000	17.330.001	17.340.000
	17.340.001	17.370.000	17.370.001	17.380.000
	17.380.001	17.390.000	17.390.001	17.400.000
	17.400.001	17.460.000	17.460.001	17.470.000
	17.470.001	17.630.000	17.630.001	17.670.000
	17.670.001	17.680.000	17.680.001	17.770.000
	17.770.001	17.780.000	17.780.001	17.810.000
	17.810.001	17.820.000	17.820.001	17.830.000
	17.830.001	17.840.000	17.840.001	17.850.000
	17.850.001	17.870.000	17.870.001	17.970.000
	17.970.001	18.010.000	18.010.001	18.020.000
	18.020.001	18.050.000	18.050.001	18.060.000
	18.060.001	18.250.000	18.250.001	18.270.000
	18.270.001	18.350.000	18.350.001	18.370.000
	18.370.001	18.380.000	18.380.001	18.390.000
	18.390.001	18.410.000	18.410.001	18.420.000
	18.420.001	18.450.000	18.450.001	18.490.000
	18.490.001	18.520.000	18.520.001	18.550.000
	18.550.001	18.560.000	18.560.001	18.660.000
	18.660.001	18.670.000	18.670.001	18.810.000
	18.810.001	18.820.000	18.820.001	18.830.000
	18.830.001	18.840.000	18.840.001	18.870.000
	18.870.001	18.890.000	18.890.001	18.910.000
	18.910.001	18.930.000	18.930.001	18.940.000
	18.940.001	18.960.000	18.960.001	19.040.000
	19.040.001	19.050.000	19.050.001	19.100.000
	19.100.001	19.110.000	19.110.001	19.150.000
	19.150.001	19.190.000	19.190.001	19.230.000
	19.230.001	19.250.000	19.250.001	19.290.000
19.290.001	19.410.000	19.410.001	19.420.000	
19.420.001	19.430.000	19.430.001	19.440.000	
19.440.001	19.460.000	19.460.001	19.490.000	
19.490.001	19.500.000	19.500.001	19.510.000	
19.510.001	19.520.000	19.520.001	19.580.000	

<i>10kb</i>	$\rho$			
	LSB		NLSB	
	Inicio	Fin	Inicio	Fin
<i>3R</i>	19.580.001	19.610.000	19.610.001	19.620.000
	19.620.001	19.640.000	19.640.001	19.660.000
	19.660.001	19.690.000	19.690.001	19.710.000
	19.710.001	19.720.000	19.720.001	19.740.000
	19.740.001	19.760.000	19.760.001	19.800.000
	19.800.001	19.830.000	19.830.001	19.850.000
	19.850.001	19.880.000	19.880.001	19.930.000
	19.930.001	19.940.000	19.940.001	19.950.000
	19.950.001	19.990.000	19.990.001	20.010.000
	20.010.001	20.020.000	20.020.001	20.040.000
	20.040.001	20.080.000	20.080.001	20.100.000
	20.100.001	20.170.000	20.170.001	20.180.000
	20.180.001	20.200.000	20.200.001	20.220.000
	20.220.001	20.230.000	20.230.001	20.350.000
	20.350.001	20.370.000	20.370.001	20.380.000
	20.380.001	20.400.000	20.400.001	20.450.000
	20.450.001	20.490.000	20.490.001	20.510.000
	20.510.001	20.550.000	20.550.001	20.580.000
	20.580.001	20.590.000	20.590.001	20.620.000
	20.620.001	20.660.000	20.660.001	20.670.000
	20.670.001	20.710.000	20.710.001	20.740.000
	20.740.001	20.780.000	20.780.001	20.840.000
	20.840.001	20.900.000	20.900.001	20.940.000
	20.940.001	20.970.000	20.970.001	21.010.000
	21.010.001	21.020.000	21.020.001	21.030.000
	21.030.001	21.170.000	21.170.001	21.180.000
	21.180.001	21.240.000	21.240.001	21.270.000
	21.270.001	21.280.000	21.280.001	21.300.000
	21.300.001	21.310.000	21.310.001	21.340.000
	21.340.001	21.350.000	21.350.001	21.420.000
	21.420.001	21.440.000	21.440.001	21.470.000
	21.470.001	21.490.000	21.490.001	21.500.000
21.500.001	21.580.000	21.580.001	21.600.000	
21.600.001	21.670.000	21.670.001	21.690.000	
21.690.001	22.020.000	22.020.001	22.030.000	

<i>10kb</i>	$\rho$			
	LSB		NLSB	
	Inicio	Fin	Inicio	Fin
<i>3R</i>	22.030.001	22.280.000	22.280.001	22.290.000
	22.290.001	22.430.000	22.430.001	22.440.000
	22.440.001	22.610.000	22.610.001	22.620.000
	22.620.001	22.680.000	22.680.001	22.770.000
	22.770.001	22.810.000	22.810.001	22.830.000
	22.830.001	22.900.000	22.900.001	22.910.000
	22.910.001	22.980.000	22.980.001	22.990.000
	22.990.001	23.040.000	23.040.001	23.050.000
	23.050.001	23.120.000	23.120.001	23.130.000
	23.130.001	23.180.000	23.180.001	23.200.000
	23.200.001	23.220.000	23.220.001	23.300.000
	23.300.001	23.330.000	23.330.001	23.340.000
	23.340.001	23.380.000	23.380.001	23.400.000
	23.400.001	23.410.000	23.410.001	23.420.000
	23.420.001	23.460.000	23.460.001	23.490.000
	23.490.001	23.510.000	23.510.001	23.520.000
	23.520.001	23.540.000	23.540.001	23.550.000
	23.550.001	23.670.000	23.670.001	23.680.000
	23.680.001	23.730.000	23.730.001	23.740.000
	23.740.001	23.790.000	23.790.001	23.810.000
	23.810.001	23.820.000	23.820.001	23.830.000
	23.830.001	23.840.000	23.840.001	23.870.000
	23.870.001	23.880.000	23.880.001	23.890.000
	23.890.001	23.900.000	23.900.001	23.910.000
	23.910.001	24.170.000	24.170.001	24.190.000
	24.190.001	24.240.000	24.240.001	24.250.000
	24.250.001	24.460.000	24.460.001	24.490.000
	24.490.001	24.530.000	24.530.001	24.540.000
	24.540.001	24.570.000	24.570.001	24.580.000
	24.580.001	24.670.000	24.670.001	24.680.000
	24.680.001	24.770.000	24.770.001	24.810.000
	24.810.001	24.820.000	24.820.001	24.840.000
24.840.001	25.000.000	25.000.001	25.040.000	
25.040.001	25.060.000	25.060.001	25.080.000	
25.080.001	25.090.000	25.090.001	25.380.000	



<i>10kb</i>	$\rho$			
	LSB		NLSB	
	Inicio	Fin	Inicio	Fin
<i>3R</i>	25.380.001	25.390.000	25.390.001	25.410.000
	25.410.001	25.440.000	25.440.001	25.450.000
	25.450.001	25.460.000	25.460.001	25.480.000
	25.480.001	25.510.000	25.510.001	25.550.000
	25.550.001	25.600.000	25.600.001	25.630.000
	25.630.001	25.670.000	25.670.001	25.680.000
	25.680.001	25.790.000	25.790.001	25.810.000
	25.810.001	25.850.000	25.850.001	25.860.000
	25.860.001	26.040.000	26.040.001	26.100.000
	26.100.001	26.110.000	26.110.001	26.140.000
	26.140.001	26.150.000	26.150.001	26.160.000
	26.160.001	26.170.000	26.170.001	26.200.000
	26.200.001	26.230.000	26.230.001	26.240.000
	26.240.001	26.410.000	26.410.001	26.420.000
	26.420.001	26.460.000	26.460.001	26.470.000
	26.470.001	26.480.000	26.480.001	26.500.000
	26.500.001	26.540.000	26.540.001	26.550.000
	26.550.001	26.560.000	26.560.001	26.570.000
	26.570.001	26.600.000	26.600.001	26.630.000
	26.630.001	26.750.000	26.750.001	26.770.000
	26.770.001	26.900.000	26.900.001	26.910.000
	26.910.001	26.920.000	26.920.001	26.940.000
	26.940.001	27.100.000	27.100.001	27.110.000
	27.110.001	27.900.000		

<i>10kb</i>	$\rho$			
	LSB		NLSB	
	Inicio	Fin	Inicio	Fin
<i>X</i>	1	270.000	270.001	280.000
	280.001	290.000	290.001	300.000
	300.001	700.000	700.001	710.000
	710.001	840.000	840.001	850.000
	850.001	880.000	880.001	900.000
	900.001	1.160.000	1.160.001	1.170.000
	1.170.001	1.190.000	1.190.001	1.200.000
	1.200.001	1.290.000	1.290.001	1.300.000
	1.300.001	1.410.000	1.410.001	1.420.000
	1.420.001	1.480.000	1.480.001	1.490.000
	1.490.001	1.530.000	1.530.001	1.540.000
	1.540.001	1.650.000	1.650.001	1.660.000
	1.660.001	1.870.000	1.870.001	1.890.000
	1.890.001	2.170.000	2.170.001	2.180.000
	2.180.001	2.200.000	2.200.001	2.310.000
	2.310.001	2.320.000	2.320.001	2.490.000
	2.490.001	2.500.000	2.500.001	2.630.000
	2.630.001	2.650.000	2.650.001	2.980.000
	2.980.001	2.990.000	2.990.001	3.060.000
	3.060.001	3.070.000	3.070.001	3.090.000
	3.090.001	3.100.000	3.100.001	3.220.000
	3.220.001	3.230.000	3.230.001	3.380.000
	3.380.001	3.410.000	3.410.001	3.500.000
	3.500.001	3.510.000	3.510.001	3.530.000
	3.530.001	3.550.000	3.550.001	3.570.000
	3.570.001	3.590.000	3.590.001	3.610.000
	3.610.001	3.640.000	3.640.001	3.650.000
	3.650.001	3.700.000	3.700.001	3.720.000
	3.720.001	3.730.000	3.730.001	3.750.000
	3.750.001	3.770.000	3.770.001	3.850.000
3.850.001	3.860.000	3.860.001	4.000.000	
4.000.001	4.020.000	4.020.001	4.060.000	
4.060.001	4.070.000	4.070.001	4.090.000	
4.090.001	4.100.000	4.100.001	4.160.000	
4.160.001	4.190.000	4.190.001	4.210.000	

<i>10kb</i>	$\rho$			
	LSB		NLSB	
	Inicio	Fin	Inicio	Fin
X	4.210.001	4.230.000	4.230.001	4.260.000
	4.260.001	4.270.000	4.270.001	4.350.000
	4.350.001	4.360.000	4.360.001	4.380.000
	4.380.001	4.390.000	4.390.001	4.500.000
	4.500.001	4.510.000	4.510.001	4.550.000
	4.550.001	4.560.000	4.560.001	4.630.000
	4.630.001	4.640.000	4.640.001	4.660.000
	4.660.001	4.670.000	4.670.001	4.720.000
	4.720.001	4.730.000	4.730.001	4.800.000
	4.800.001	4.830.000	4.830.001	5.030.000
	5.030.001	5.050.000	5.050.001	5.080.000
	5.080.001	5.090.000	5.090.001	5.100.000
	5.100.001	5.110.000	5.110.001	5.160.000
	5.160.001	5.170.000	5.170.001	5.190.000
	5.190.001	5.210.000	5.210.001	5.240.000
	5.240.001	5.250.000	5.250.001	5.330.000
	5.330.001	5.370.000	5.370.001	5.380.000
	5.380.001	5.410.000	5.410.001	5.450.000
	5.450.001	5.470.000	5.470.001	5.590.000
	5.590.001	5.600.000	5.600.001	5.620.000
	5.620.001	5.660.000	5.660.001	5.720.000
	5.720.001	5.750.000	5.750.001	5.790.000
	5.790.001	5.800.000	5.800.001	5.930.000
	5.930.001	5.940.000	5.940.001	6.040.000
	6.040.001	6.050.000	6.050.001	6.200.000
	6.200.001	6.210.000	6.210.001	6.270.000
	6.270.001	6.280.000	6.280.001	6.380.000
	6.380.001	6.410.000	6.410.001	6.560.000
	6.560.001	6.580.000	6.580.001	6.650.000
	6.650.001	6.670.000	6.670.001	6.750.000
6.750.001	6.760.000	6.760.001	6.780.000	
6.780.001	6.790.000	6.790.001	6.800.000	
6.800.001	6.820.000	6.820.001	6.880.000	
6.880.001	6.890.000	6.890.001	6.930.000	
6.930.001	6.940.000	6.940.001	6.950.000	

10kb	$\rho$			
	LSB		NLSB	
	Inicio	Fin	Inicio	Fin
X	6.950.001	6.970.000	6.970.001	6.980.000
	6.980.001	7.000.000	7.000.001	7.010.000
	7.010.001	7.020.000	7.020.001	7.030.000
	7.030.001	7.050.000	7.050.001	7.090.000
	7.090.001	7.100.000	7.100.001	7.110.000
	7.110.001	7.120.000	7.120.001	7.130.000
	7.130.001	7.140.000	7.140.001	7.150.000
	7.150.001	7.160.000	7.160.001	7.210.000
	7.210.001	7.220.000	7.220.001	7.250.000
	7.250.001	7.260.000	7.260.001	7.290.000
	7.290.001	7.310.000	7.310.001	7.330.000
	7.330.001	7.340.000	7.340.001	7.380.000
	7.380.001	7.390.000	7.390.001	7.480.000
	7.480.001	7.530.000	7.530.001	7.540.000
	7.540.001	7.570.000	7.570.001	7.590.000
	7.590.001	7.610.000	7.610.001	7.620.000
	7.620.001	7.630.000	7.630.001	7.640.000
	7.640.001	7.650.000	7.650.001	7.700.000
	7.700.001	7.710.000	7.710.001	7.760.000
	7.760.001	7.770.000	7.770.001	7.800.000
	7.800.001	7.850.000	7.850.001	7.940.000
	7.940.001	7.960.000	7.960.001	8.010.000
	8.010.001	8.020.000	8.020.001	8.110.000
	8.110.001	8.120.000	8.120.001	8.310.000
	8.310.001	8.320.000	8.320.001	8.340.000
	8.340.001	8.350.000	8.350.001	8.360.000
	8.360.001	8.370.000	8.370.001	8.410.000
	8.410.001	8.420.000	8.420.001	8.430.000
	8.430.001	8.440.000	8.440.001	8.450.000
	8.450.001	8.460.000	8.460.001	8.470.000
8.470.001	8.480.000	8.480.001	8.530.000	
8.530.001	8.540.000	8.540.001	8.600.000	
8.600.001	8.610.000	8.610.001	8.630.000	
8.630.001	8.640.000	8.640.001	8.660.000	
8.660.001	8.670.000	8.670.001	8.680.000	

10kb	$\rho$			
	LSB		NLSB	
	Inicio	Fin	Inicio	Fin
X	8.680.001	8.690.000	8.690.001	8.740.000
	8.740.001	8.750.000	8.750.001	8.760.000
	8.760.001	8.780.000	8.780.001	8.790.000
	8.790.001	8.800.000	8.800.001	8.810.000
	8.810.001	8.820.000	8.820.001	8.830.000
	8.830.001	8.840.000	8.840.001	8.920.000
	8.920.001	8.930.000	8.930.001	8.990.000
	8.990.001	9.000.000	9.000.001	9.050.000
	9.050.001	9.060.000	9.060.001	9.140.000
	9.140.001	9.150.000	9.150.001	9.430.000
	9.430.001	9.440.000	9.440.001	9.520.000
	9.520.001	9.530.000	9.530.001	9.540.000
	9.540.001	9.550.000	9.550.001	9.610.000
	9.610.001	9.620.000	9.620.001	9.810.000
	9.810.001	9.820.000	9.820.001	9.990.000
	9.990.001	10.000.000	1.0.00.0.001	10.100.000
	10.100.001	10.110.000	10.110.001	10.160.000
	10.160.001	10.170.000	10.170.001	10.340.000
	10.340.001	10.350.000	10.350.001	10.360.000
	10.360.001	10.390.000	10.390.001	10.420.000
	10.420.001	10.430.000	10.430.001	10.450.000
	10.450.001	10.460.000	10.460.001	10.470.000
	10.470.001	10.480.000	10.480.001	10.500.000
	10.500.001	10.540.000	10.540.001	10.550.000
	10.550.001	10.560.000	10.560.001	10.630.000
	10.630.001	10.640.000	10.640.001	10.650.000
	10.650.001	10.660.000	10.660.001	10.690.000
	10.690.001	10.700.000	10.700.001	10.830.000
	10.830.001	10.840.000	10.840.001	10.870.000
	10.870.001	10.880.000	10.880.001	10.940.000
10.940.001	10.950.000	10.950.001	10.990.000	
10.990.001	11.000.000	11.000.001	11.190.000	
11.190.001	11.200.000	11.200.001	11.250.000	
11.250.001	11.270.000	11.270.001	11.290.000	
11.290.001	11.300.000	11.300.001	11.310.000	

<i>10kb</i>	$\rho$			
	LSB		NLSB	
	Inicio	Fin	Inicio	Fin
X	11.310.001	11.340.000	11.340.001	11.360.000
	11.360.001	11.370.000	11.370.001	11.380.000
	11.380.001	11.400.000	11.400.001	11.440.000
	11.440.001	11.480.000	11.480.001	11.600.000
	11.600.001	11.620.000	11.620.001	11.640.000
	11.640.001	11.650.000	11.650.001	11.660.000
	11.660.001	11.690.000	11.690.001	11.780.000
	11.780.001	11.790.000	11.790.001	11.800.000
	11.800.001	11.810.000	11.810.001	11.960.000
	11.960.001	11.970.000	11.970.001	12.020.000
	12.020.001	12.040.000	12.040.001	12.460.000
	12.460.001	12.480.000	12.480.001	12.580.000
	12.580.001	12.610.000	12.610.001	12.680.000
	12.680.001	12.690.000	12.690.001	12.920.000
	12.920.001	12.930.000	12.930.001	12.940.000
	12.940.001	12.950.000	12.950.001	13.070.000
	13.070.001	13.080.000	13.080.001	13.170.000
	13.170.001	13.180.000	13.180.001	13.330.000
	13.330.001	13.340.000	13.340.001	13.420.000
	13.420.001	13.430.000	13.430.001	13.500.000
	13.500.001	13.510.000	13.510.001	13.540.000
	13.540.001	13.560.000	13.560.001	13.580.000
	13.580.001	13.590.000	13.590.001	13.710.000
	13.710.001	13.720.000	13.720.001	13.730.000
	13.730.001	13.740.000	13.740.001	13.820.000
	13.820.001	13.830.000	13.830.001	13.860.000
	13.860.001	13.870.000	13.870.001	13.990.000
	13.990.001	14.010.000	14.010.001	14.130.000
	14.130.001	14.140.000	14.140.001	14.170.000
	14.170.001	14.180.000	14.180.001	14.190.000
14.190.001	14.230.000	14.230.001	14.470.000	
14.470.001	14.480.000	14.480.001	14.710.000	
14.710.001	14.720.000	14.720.001	14.730.000	
14.730.001	14.740.000	14.740.001	14.770.000	
14.770.001	14.780.000	14.780.001	14.830.000	

<i>10kb</i>	$\rho$			
	LSB		NLSB	
	Inicio	Fin	Inicio	Fin
X	14.830.001	14.880.000	14.880.001	14.890.000
	14.890.001	14.930.000	14.930.001	14.960.000
	14.960.001	15.010.000	15.010.001	15.050.000
	15.050.001	15.060.000	15.060.001	15.210.000
	15.210.001	15.240.000	15.240.001	15.410.000
	15.410.001	15.420.000	15.420.001	15.530.000
	15.530.001	15.540.000	15.540.001	15.560.000
	15.560.001	15.570.000	15.570.001	15.590.000
	15.590.001	15.610.000	15.610.001	15.760.000
	15.760.001	15.770.000	15.770.001	15.810.000
	15.810.001	15.820.000	15.820.001	15.850.000
	15.850.001	15.870.000	15.870.001	15.900.000
	15.900.001	15.950.000	15.950.001	15.990.000
	15.990.001	16.030.000	16.030.001	16.070.000
	16.070.001	16.080.000	16.080.001	16.090.000
	16.090.001	16.100.000	16.100.001	16.130.000
	16.130.001	16.150.000	16.150.001	16.170.000
	16.170.001	16.180.000	16.180.001	16.190.000
	16.190.001	16.200.000	16.200.001	16.210.000
	16.210.001	16.220.000	16.220.001	16.230.000
	16.230.001	16.240.000	16.240.001	16.300.000
	16.300.001	16.310.000	16.310.001	16.320.000
	16.320.001	16.360.000	16.360.001	16.480.000
	16.480.001	16.490.000	16.490.001	16.680.000
	16.680.001	16.690.000	16.690.001	16.830.000
	16.830.001	16.840.000	16.840.001	17.000.000
	17.000.001	17.010.000	17.010.001	17.030.000
	17.030.001	17.050.000	17.050.001	17.080.000
	17.080.001	17.090.000	17.090.001	17.120.000
	17.120.001	17.130.000	17.130.001	17.190.000
17.190.001	17.200.000	17.200.001	17.270.000	
17.270.001	17.280.000	17.280.001	17.290.000	
17.290.001	17.300.000	17.300.001	17.400.000	
17.400.001	17.410.000	17.410.001	17.550.000	
17.550.001	17.560.000	17.560.001	17.690.000	

<i>10kb</i>	$\rho$			
	LSB		NLSB	
	Inicio	Fin	Inicio	Fin
X	17.690.001	17.700.000	17.700.001	17.730.000
	17.730.001	17.760.000	17.760.001	17.780.000
	17.780.001	17.800.000	17.800.001	17.920.000
	17.920.001	18.020.000	18.020.001	18.030.000
	18.030.001	18.040.000	18.040.001	18.080.000
	18.080.001	18.110.000	18.110.001	18.120.000
	18.120.001	18.150.000	18.150.001	18.160.000
	18.160.001	18.170.000	18.170.001	18.190.000
	18.190.001	18.220.000	18.220.001	18.230.000
	18.230.001	18.240.000	18.240.001	18.260.000
	18.260.001	18.270.000	18.270.001	18.330.000
	18.330.001	18.340.000	18.340.001	18.480.000
	18.480.001	18.490.000	18.490.001	18.500.000
	18.500.001	18.530.000	18.530.001	18.540.000
	18.540.001	18.610.000	18.610.001	18.670.000
	18.670.001	18.690.000	18.690.001	18.830.000
	18.830.001	18.840.000	18.840.001	18.850.000
	18.850.001	18.860.000	18.860.001	19.030.000
	19.030.001	19.040.000	19.040.001	19.170.000
	19.170.001	19.190.000	19.190.001	19.220.000
	19.220.001	19.240.000	19.240.001	19.270.000
	19.270.001	19.280.000	19.280.001	19.300.000
	19.300.001	19.310.000	19.310.001	19.340.000
	19.340.001	19.350.000	19.350.001	19.400.000
	19.400.001	19.410.000	19.410.001	19.460.000
	19.460.001	19.480.000	19.480.001	19.520.000
	19.520.001	19.530.000	19.530.001	19.550.000
	19.550.001	19.560.000	19.560.001	19.760.000
	19.760.001	19.770.000	19.770.001	19.790.000
	19.790.001	19.820.000	19.820.001	19.830.000
19.830.001	19.860.000	19.860.001	19.880.000	
19.880.001	19.890.000	19.890.001	20.040.000	
20.040.001	20.050.000	20.050.001	20.070.000	
20.070.001	20.080.000	20.080.001	20.090.000	
20.090.001	20.110.000	20.110.001	20.120.000	



<i>10kb</i>	$\rho$			
	LSB		NLSB	
	Inicio	Fin	Inicio	Fin
<i>X</i>	20.120.001	20.130.000	20.130.001	20.140.000
	20.140.001	20.150.000	20.150.001	20.170.000
	20.170.001	20.180.000	20.180.001	20.260.000
	20.260.001	20.280.000	20.280.001	20.290.000
	20.290.001	20.310.000	20.310.001	20.580.000
	20.580.001	20.620.000	20.620.001	20.630.000
	20.630.001	20.640.000	20.640.001	20.650.000
	20.650.001	20.770.000	20.770.001	20.780.000
	20.780.001	20.810.000	20.810.001	20.830.000
	20.830.001	20.840.000	20.840.001	20.870.000
	20.870.001	20.880.000	20.880.001	20.910.000
	20.910.001	20.920.000	20.920.001	20.980.000
	20.980.001	21.050.000	21.050.001	21.160.000
	21.160.001	21.210.000	21.210.001	21.220.000
	21.220.001	21.270.000	21.270.001	21.330.000
	21.330.001	21.340.000	21.340.001	21.360.000
	21.360.001	21.370.000	21.370.001	21.380.000
	21.380.001	21.450.000	21.450.001	21.460.000
	21.460.001	21.680.000	21.680.001	21.690.000
	21.690.001	21.850.000	21.850.001	21.950.000
21.950.001	22.140.000	22.140.001	22.150.000	
22.150.001	22.290.000	22.290.001	22.300.000	
22.300.001	22.310.000	22.310.001	22.320.000	
22.320.001	22.420.000			

Mapa de coordenadas para  $c$  y  $\rho$  en ventanas no solapantes de 1Mb, 100kb y 10kb.

**Tabla suplementaria 3. Regiones centroméricas y teloméricas**

**a) 100kb**

<b>Cromosoma</b>	<b>Centrómero</b>	<b>Telómero</b>	<b>Central</b>
2L	Inicio: 16.600.001 Fin: 23.100.000	Inicio: 1 Fin: 800.000	Inicio: 800.001 Fin: 16.600.000
2R	Inicio: 1 Fin: 6.200.000	Inicio: 19.800.001 Fin: 21.200.000	Inicio: 6.200.001 Fin: 19.800.000
3L	Inicio: 18.300.001 Fin: 24.600.000	Inicio: 1 Fin: 1.000.000	Inicio: 1.000.001 Fin: 18.300.000
3R	Inicio: 1 Fin: 9.500.000	Inicio: 26.700.001 Fin: 28.000.000	Inicio: 9.500.001 Fin: 26.700.000
X	Inicio: 21.100.001 Fin: 22.500.000	Inicio: 1 Fin: 2.200.000	Inicio: 2.200.001 Fin: 21.100.000

**b) 10kb**

<b>Cromosoma</b>	<b>Centrómero</b>	<b>Telómero</b>	<b>Central</b>
2L	Inicio: 21.640.001 Fin: 22.500.000	Inicio: 1 Fin: 500.000	Inicio: 500.001 Fin: 21.640.000
2R	Inicio: 1 Fin: 340.000	Inicio: 19.970.001 Fin: 21.140.000	Inicio: 340.001 Fin: 19.970.000
3L	Inicio: 24.140.001 Fin: 24.540.000	Inicio: 1 Fin: 110.000	Inicio: 100.001 Fin: 24.140.000
3R	Inicio: 1 Fin: 650.000	Inicio: 27.110.001 Fin: 27.900.000	Inicio: 650.001 Fin: 27.110.000
X	Inicio: 22.320.001 Fin: 22.420.000	Inicio: 1 Fin: 270.000	Inicio: 270.001 Fin: 22.320.000

Regiones consideradas como centroméricas y teloméricas. Nótese que es distinto a las regiones centroméricas y teloméricas usadas en el apartado de resultados 3.1 y 3.2 para referirse a un tercio del cromosoma.

**Tabla suplementaria 4.** Pruebas de significación para la existencia de diferencias entre brazos cromosómicos de las estimas de selección

$\alpha$		Kruskal-Wallis $\chi^2$	Grados de libertad	valor p	
	<i>Diferencias entre cromosomas</i>	45,3436	4	$3,37 \cdot 10^{-9}$	
	<i>Diferencias entre autosomas</i>	28,3661	3	$3,04 \cdot 10^{-6}$	
$\alpha$	2L	2R	3L	3R	X
2L	-----	-----	-----	-----	-----
2R	0,1638	-----	-----	-----	-----
3L	0,0130	0,2991	-----	-----	-----
3R	$5,57 \cdot 10^{-7}$	$2,48 \cdot 10^{-4}$	$8,69 \cdot 10^{-3}$	-----	-----
X	0,2202	0,0138	$5,86 \cdot 10^{-4}$	$3,53 \cdot 10^{-9}$	-----

$k_a/k_s$		Kruskal-Wallis $\chi^2$	Grados de libertad	valor p	
	<i>Diferencias entre cromosomas</i>	15,8616	4	$3,21 \cdot 10^{-3}$	
	<i>Diferencias entre autosomas</i>	8,9472	3	$3,00 \cdot 10^{-2}$	
$k_a/k_s$	2L	2R	3L	3R	X
2L	-----	-----	-----	-----	-----
2R	0,7872	-----	-----	-----	-----
3L	0,0117	0,0238	-----	-----	-----
3R	0,0638	0,1125	0,4495	-----	-----
X	0,3836	0,2550	$7,39 \cdot 10^{-4}$	$4,89 \cdot 10^{-3}$	-----

$\omega_a$		Kruskal-Wallis $\chi^2$	Grados de libertad	valor p	
	<i>Diferencias entre cromosomas</i>	41,3923	4	$2,23 \cdot 10^{-8}$	
	<i>Diferencias entre autosomas</i>	25,6478	3	$1,13 \cdot 10^{-5}$	
$\omega_a$	2L	2R	3L	3R	X
2L	-----	-----	-----	-----	-----
2R	0,1756	-----	-----	-----	-----
3L	$1,44 \cdot 10^{-3}$	0,0779	-----	-----	-----
3R	$4,24 \cdot 10^{-6}$	$1,02 \cdot 10^{-3}$	0,0806	-----	-----
X	0,3528	0,0284	$8,90 \cdot 10^{-5}$	$6,59 \cdot 10^{-8}$	-----

$\omega_d$	Kruskal-Wallis $\chi^2$	Grados de libertad	valor p		
<i>Diferencias entre cromosomas</i>	23,6313	4	$9,47 \cdot 10^{-5}$		
<i>Diferencias entre autosomas</i>	18,4594	3	$3,54 \cdot 10^{-4}$		
$\omega_d$	2L	2R	3L	3R	X
2L	-----	-----	-----	-----	-----
2R	0,2777	-----	-----	-----	-----
3L	0,5069	0,6473	-----	-----	-----
3R	$9,54 \cdot 10^{-5}$	$7,13 \cdot 10^{-3}$	$1,12 \cdot 10^{-3}$	-----	-----
X	0,6347	0,1251	0,2985	$3,77 \cdot 10^{-5}$	-----

Prueba de Wilcoxon y Kruskal-Wallis entre los brazos cromosómicos para  $\alpha$ ,  $\omega_a$ ,  $\omega_d$  y  $k_a/k_s$ .

## 7.2. Anexo A: Información básica de la secuenciación de las 158 líneas DGRP

Línea DGRP	Cobertura (x)	Longitud de las lecturas	Línea DGRP	Cobertura (x)	Longitud de las lecturas
<i>RAL-21</i>	19,80	95	<i>RAL-405</i>	25,15	95
<i>RAL-26</i>	18,70	75	<i>RAL-406</i>	30,30	95
<i>RAL-28</i>	18,10	75	<i>RAL-409</i>	12,65	125
<i>RAL-38</i>	34,10	95	<i>RAL-426</i>	25,74	95
<i>RAL-40</i>	41,27	95	<i>RAL-427</i>	20,61	45
<i>RAL-41</i>	16,76	75	<i>RAL-437</i>	13,93	125
<i>RAL-42</i>	24,27	95	<i>RAL-439</i>	21,44	95
<i>RAL-45</i>	18,35	75	<i>RAL-440</i>	20,80	95
<i>RAL-49</i>	19,65	75	<i>RAL-441</i>	22,88	95
<i>RAL-57</i>	38,28	100	<i>RAL-443</i>	33,35	95
<i>RAL-59</i>	19,57	75	<i>RAL-461</i>	24,45	95
<i>RAL-69</i>	17,26	75	<i>RAL-491</i>	18,65	75
<i>RAL-73</i>	18,12	75	<i>RAL-492</i>	27,07	95
<i>RAL-75</i>	22,31	110	<i>RAL-502</i>	26,88	95
<i>RAL-83</i>	20,50	75	<i>RAL-508</i>	26,19	95
<i>RAL-85</i>	18,54	75	<i>RAL-509</i>	19,12	75
<i>RAL-88</i>	16,80	75	<i>RAL-513</i>	23,48	95
<i>RAL-91</i>	17,07	75	<i>RAL-517</i>	45,97	75
<i>RAL-93</i>	18,45	75	<i>RAL-531</i>	22,11	95
<i>RAL-101</i>	15,19	75	<i>RAL-535</i>	18,86	75
<i>RAL-105</i>	16,95	75	<i>RAL-555</i>	23,74	75
<i>RAL-109</i>	13,42	125	<i>RAL-563</i>	17,95	75
<i>RAL-129</i>	13,85	75	<i>RAL-589</i>	16,32	100
<i>RAL-136</i>	18,12	75	<i>RAL-591</i>	15,82	75
<i>RAL-138</i>	34,00	100	<i>RAL-595</i>	15,94	75
<i>RAL-142</i>	23,85	110	<i>RAL-639</i>	20,71	75
<i>RAL-149</i>	17,45	75	<i>RAL-642</i>	17,42	75
<i>RAL-153</i>	12,54	75	<i>RAL-646</i>	15,52	75
<i>RAL-158</i>	14,82	75	<i>RAL-703</i>	17,72	95
<i>RAL-161</i>	17,72	75	<i>RAL-705</i>	20,60	75
<i>RAL-176</i>	38,48	75	<i>RAL-707</i>	22,06	75
<i>RAL-177</i>	29,77	95	<i>RAL-712</i>	20,27	75
<i>RAL-181</i>	30,66	75	<i>RAL-714</i>	21,34	75
<i>RAL-195</i>	18,52	75	<i>RAL-716</i>	15,74	75
<i>RAL-208</i>	34,51	64	<i>RAL-721</i>	15,82	75

<i>RAL-217</i>	16,68	75	<i>RAL-727</i>	33,64	75
<i>RAL-227</i>	15,03	75	<i>RAL-730</i>	17,20	75
<i>RAL-228</i>	15,13	75	<i>RAL-732</i>	20,47	75
<i>RAL-229</i>	13,49	125	<i>RAL-737</i>	28,13	75
<i>RAL-233</i>	15,71	75	<i>RAL-738</i>	31,74	75
<i>RAL-235</i>	22,61	95	<i>RAL-757</i>	32,87	75
<i>RAL-237</i>	20,33	75	<i>RAL-761</i>	18,74	75
<i>RAL-239</i>	18,14	75	<i>RAL-765</i>	20,48	75
<i>RAL-256</i>	17,98	75	<i>RAL-774</i>	13,75	125
<i>RAL-280</i>	18,00	75	<i>RAL-776</i>	19,39	75
<i>RAL-287</i>	11,67	75	<i>RAL-783</i>	16,89	75
<i>RAL-309</i>	16,53	75	<i>RAL-786</i>	14,73	45
<i>RAL-310</i>	15,88	75	<i>RAL-787</i>	19,17	75
<i>RAL-313</i>	15,02	45	<i>RAL-790</i>	21,03	95
<i>RAL-317</i>	12,42	75	<i>RAL-796</i>	18,89	75
<i>RAL-318</i>	19,03	75	<i>RAL-799</i>	20,83	75
<i>RAL-320</i>	29,42	95	<i>RAL-801</i>	14,79	75
<i>RAL-321</i>	41,75	95	<i>RAL-802</i>	13,96	125
<i>RAL-325</i>	14,93	75	<i>RAL-804</i>	18,73	125
<i>RAL-332</i>	31,53	75	<i>RAL-805</i>	20,26	75
<i>RAL-338</i>	14,20	75	<i>RAL-808</i>	15,39	75
<i>RAL-350</i>	13,23	75	<i>RAL-810</i>	19,41	125
<i>RAL-352</i>	19,18	75	<i>RAL-812</i>	20,44	75
<i>RAL-356</i>	18,30	75	<i>RAL-818</i>	18,56	75
<i>RAL-357</i>	27,21	45	<i>RAL-820</i>	18,16	75
<i>RAL-358</i>	19,40	75	<i>RAL-822</i>	20,86	110
<i>RAL-359</i>	24,84	95	<i>RAL-832</i>	17,33	95
<i>RAL-362</i>	12,52	75	<i>RAL-837</i>	24,58	95
<i>RAL-365</i>	19,75	75	<i>RAL-852</i>	40,42	75
<i>RAL-367</i>	16,02	75	<i>RAL-855</i>	23,01	110
<i>RAL-370</i>	25,45	95	<i>RAL-857</i>	25,10	110
<i>RAL-371</i>	14,50	85	<i>RAL-859</i>	18,28	75
<i>RAL-373</i>	25,45	125	<i>RAL-861</i>	15,14	75
<i>RAL-374</i>	16,63	95	<i>RAL-879</i>	14,72	75
<i>RAL-375</i>	41,91	75	<i>RAL-882</i>	21,73	75
<i>RAL-377</i>	27,04	95	<i>RAL-884</i>	16,63	75
<i>RAL-379</i>	14,64	75	<i>RAL-887</i>	23,24	95
<i>RAL-380</i>	36,73	74	<i>RAL-890</i>	19,86	75

---

<i>RAL-381</i>	26,06	75
<i>RAL-383</i>	23,59	95
<i>RAL-386</i>	14,43	75
<i>RAL-391</i>	47,62	75
<i>RAL-392</i>	27,58	95
<i>RAL-399</i>	19,61	75

<i>RAL-892</i>	25,43	95
<i>RAL-894</i>	20,56	95
<i>RAL-897</i>	32,81	75
<i>RAL-907</i>	21,37	95
<i>RAL-908</i>	24,38	95
<i>RAL-911</i>	17,19	95

---

## 7.3. Anexo B: Programas Perl

Código desarrollado para la estimación del polimorfismo y la divergencia en las distintas clases funcionales de sitios de las secuencias génicas de los genomas DGRP.

### 7.3.1. Fragmentar los alineamientos cromosómicos en alineamientos génicos

```
#!/usr/bin/perl
use warnings;
use strict;

my $brazoCromosoma = ("2R");
my $fragmento = 'CDS';
my $freeze =
"..../freezel.consensus/noDuplicates/$brazoCromosoma.alloutgroups.nodupl.fasta";
my @trans_pn_ini_fin;
my $asdf;
my $archivoLectura;
my $archivoEscritura;

open($archivoLectura, "<", "Anotaciones/dmel~$brazoCromosoma~r5.13-$fragmento.gff") or
die "Error 1";
print "Leyendo archivo: dmel~$brazoCromosoma~r5.13-$fragmento.gff\n";
my $i = 0;
while(<$archivoLectura>)
{
    $_ =~ s/\n//g;
    my @linea = split(/\t/, $_);
    if ($#linea >= 8 and $linea[2] eq "CDS")
    {
        my $pn_ini_fin = "$linea[6]\t$linea[3]\t$linea[4]";
        @linea = split("Parent=", $linea[8]);
        $linea[1] =~ s/\n//g;
        $linea[1] =~ s/\s//g;
        foreach my $trans (split(",", $linea[1]))
        {
            $trans_pn_ini_fin[$i] = "$trans\t$pn_ini_fin";
            $i++;
        }
    }
}
close $archivoLectura;

#####

#print "Ordenando...\n";
@trans_pn_ini_fin = sort @trans_pn_ini_fin;

#####

foreach my $trans_pn_ini_fin (@trans_pn_ini_fin)
{
    #print PRUEBA $trans_pn_ini_fin."\n";
    #$trans_pn_ini_fin =~ s/\n//g;
    my @aux = split(/\t/, $trans_pn_ini_fin);
    my $ini = $aux[2]-1;
    my $length = $aux[3]-$ini;
    #secuecnias
    open ($archivoLectura, "<", $freeze) or die "Error dsim";
    #print "Leyendo archivo: $brazoCromosoma.all.dsim.freezel.fasta\n";
    open (ESCRITURA,
">exon/2L/all/$brazoCromosoma"."_"."$aux[0]"."_"."$aux[2]:$aux[3].fasta") or die "Error
3";
    print "Guardando resultados en: $brazoCromosoma"."_"."$aux[0].txt\n";
    while (<$archivoLectura>)
    {
        #my $seq = "";
        $_ =~ s/\n//g;
        $_ =~ s/>/_/g;
        my $line = $_;
        $_ = <$archivoLectura>;
        my $seq = (substr $_, $ini, $length);
    #    $i = 2;
    #    while ($aux[$i])
```



```

# {
#   my $ini = $aux[$i]-1;
#   my $length = $aux[$i+1]-$ini;
#   $seq = $seq . (substr $_, $ini, $length);
#   $i = $i+2;
# }
#   if ($aux[1] eq "-")
#   {
#     print "La secuencia es negativa => Reversando...\n";
#     $seq = reverse $seq;
#     $seq =~tr/ATCG/TAGC/;
#   }

# Para el DNAsp
#   $seq =~s/[^ATCG-]/N/g;
#   print ESCRITURA ">$aux[0]$line\_ $aux[2]-$aux[3]\n";
#   print ESCRITURA "$seq\n\n";

# Quitar N's
#   unless ($seq =~/N/)
#   {
#     print ESCRITURA ">$aux[0]$line\_ $aux[2]-$aux[3]\n";
#     print ESCRITURA "$seq\n\n";
#   }
# }
# close ESCRITURA;
# close $archivoLectura;
# $asdf = <STDIN>;
# }

#print "Fin del archivo: dmel\-$brazoCromosoma\r5.13-fragmento.gff\n\n\n";

print "FIN!!!";

```

### 7.3.2. Eliminar sitios ambiguos

```

#!/usr/bin/perl
opendir ( DIR, "/home/dgrp/Maite/Perl_Feature/2L/TE" ) || die "Error in opening dir
$dirname-startN\n";
while( ($filename = readdir(DIR))
{
  chomp $filename;
  next if ($filename !~/^[^_]+_[^\.] +\.txt$/);
  print("$filename\n");

  $exec_command = "perl DGRP_N.pl ".$filename;
  system($exec_command);
} # final del bucle readdir
closedir(DIR);

#Entrada programa general
#Crear archivo salida
$filename = $ARGV[0];
($id) = ($filename =~/^[^\.] +/);

open(FILE, "TE/$filename")or die "Error 1";
open (RESULTS, ">TE_without_N/$id.fasta")or die "Error 2";

@fastas = <FILE>;
$fasta = join(' ', @fastas);
$fasta =~s/\s/\t/g;
$fasta =~s/>/\n/g;
$fasta =~s/Line//g;
$fasta =~s/2R/2r/g;
@poly_fasta_N = split ("\n",$fasta);

$vacio =shift @poly_fasta_N;
$recoded =shift @poly_fasta_N;
$ere =shift @poly_fasta_N;
$yak =shift @poly_fasta_N;
$sec =shift @poly_fasta_N;
$sim =shift @poly_fasta_N;

#Eliminar N's
foreach $fasta_N (@poly_fasta_N)
{

```

```

if ($fasta_N =~/^FB/)
{
# $fasta_N =~s/[^ATCG-]/N/g;
# print RESULTS $fasta_N."\n";
unless ($fasta_N =~/[NMRWSYK]/)
{
    $poly_fasta = $fasta_N;
    push(@poly_fasta, $poly_fasta);
}
}
}
print "Eliminando Ns...";
#print RESULTS @poly_fasta;
$recoded =~s/\t/\n/g;
$ere =~s/\t/\n/g;
$yak =~s/\t/\n/g;
$sec =~s/\t/\n/g;
$sim =~s/\t/\n/g;

print RESULTS ">$recoded";
print RESULTS ">$ere";
print RESULTS ">$yak";
print RESULTS ">$sec";
print RESULTS ">$sim";

foreach $poly_fasta (@poly_fasta)
{
    $poly_fasta =~s/\t/\n/g;
    print RESULTS ">$poly_fasta";
}

```

### 7.3.3. Estimaciones de polimorfismo utilizando PDA

```

#!/usr/bin/perl

#####
##### SynNoSyn PDA #####
#####

use SynNoSyn_tocado;

open(RESULTS, ">TEdiv_dyak_2L.txt");
print RESULTS "chr FBtr num_stop_codons C Sites promig_S Ks_promig Ks_JC_promig
Ks_JC_Var_promig promig_N Ka_promig Ka_JC_promig Ka_JC_Var_promig\n";
close RESULTS;

opendir ( DIR, "/home/dgrp/Maite/Perl_Feature/2L/TE_dyak" ) || die "Error in opening dir
$dirname\n";
while(($filename = readdir(DIR))
{
    chomp $filename;
    next if ($filename !~/^[^_]+_[^\.]\.fasta$/);
    print("$filename\n");

    $exec_command = "perl DGRP_pda.pl ".$filename;
    system($exec_command);
} # final del bucle readdir
closedir(DIR);

use SynNoSyn_tocado;

open(RESULTS, ">>TEdiv_dyak_2L.txt");
$filename = $ARGV[0];
(chr, $FBtr) = ($filename =~/^[^_]+_[^\.]\.fasta$/);

open(FILE, "TE_dyak/".$filename) or die "No puedo abrir los archivos";

#Mi forma de obtener secuencias
while (<FILE>)
{
    if ($_ =~/^>/)
    {
        $poly_def = $_;
        push(@poly_def, $poly_def);
    } elsif ($_ =~/[AGCT]/){
        $poly_seq = $_;
    }
}

```

```

    push(@poly_seq, $poly_seq);
  }
}
shift(@poly_def); #Para eliminar la secuencia de vacio
shift(@poly_seq); #Para eliminar la secuencia de vacio
shift(@poly_def); #Para eliminar la secuencia de recoded
#shift(@poly_seq); #Para eliminar la secuencia de recoded
#shift(@poly_def); #Para eliminar la secuencia de dere
#shift(@poly_seq); #Para eliminar la secuencia de dere
#shift(@poly_def); #Para eliminar la secuencia de dyak
#shift(@poly_seq); #Para eliminar la secuencia de dyak
#shift(@poly_def); #Para eliminar la secuencia de dsec
#shift(@poly_seq); #Para eliminar la secuencia de dsec
#shift(@poly_def); #Para eliminar la secuencia de dsim
#shift(@poly_seq); #Para eliminar la secuencia de dsim

$numero_secuencias = @poly_seq; #numero de secuencias por alineamiento sin Ns.

my($num_stop_codons, $C, $Sites, $promig_S, $Ks_promig, $Ks_JC_promig,
$Ks_JC_Var_promig, $promig_N, $Ka_promig,
$Ka_JC_promig, $Ka_JC_Var_promig, $seques2, $SSS, $NS, $seques, $SilentDif, $NSynDif,
$SilentPos, $NSynPos, $Ks, $Ka,
$Ks_JC, $Ka_JC, $Ks_JC_Var, $Ka_JC_Var) = SynNoSyn_tocado-
>SynNoSyn_Calculate(\@poly_seq, \@poly_def);

print RESULTS "$chr $FBtr $numero_secuencias $num_stop_codons $C $Sites $promig_S
$Ks_promig $Ks_JC_promig $Ks_JC_Var_promig $promig_N $Ka_promig $Ka_JC_promig
$Ka_JC_Var_promig\n";

```

### 7.3.4. Estimaciones de divergencia utilizando PDA

```

#!/usr/bin/perl
opendir ( DIR, "/home/dgrp/Maite/Perl_Feature/2L/TE_without_N" ) || die "Error in
opening dir $dirname\n";
while( ($filename = readdir(DIR))
{
  chomp $filename;
  next if ($filename !~/^[^_]+_[^\.]\.fasta$/);
  print("$filename\n");
  $exec_command = "perl DGRP_divergence.pl ".$filename;
  system($exec_command);
} # final del bucle readdir
closedir(DIR);

#Obtención de dos secuencias para calcular divergencia con PDA.
#Secuencia divergente y secuencia polimórfica consenso de todas las #lineas.

#!/usr/bin/perl

use consensus;
use vars qw(@ISA $NO_CONSENSUS_SYMBOL);
#use Bio::UnivAln;
use UnivAln;

@ISA = qw(Bio::UnivAln);
$NO_CONSENSUS_SYMBOL = $Bio::UnivAln::_NO_CONSENSUS_SYMBOL;

#Entrada servidos:
#Crear archivo salida
my $filename = $ARGV[0];
(my $id) = ($filename =~/^[^\.]+/);

open(FILE, "TE_without_N/".$filename)or die "Error 1";
open(RESULTS, ">TE_dere/$id.fasta")or die "Error 2";

#Cada id+nombre un elemento del array
my @fastas = <FILE>;
my $fasta_1 = join(' ', @fastas);
$fasta_1 =~s/\s/\t/g;
$fasta_1 =~s/>/\n/g;
my @poly_fastas = split ("\n",$fasta_1);
#Borrar la primera linea que esta vacia debido a que los archivos siempre empiezan por
>.
shift @poly_fastas;

#Retirar la secuencia del outgroup

```

```

$recoded = shift @poly_fasta;
$dere = shift @poly_fasta;
$dyak = shift @poly_fasta;
$dsec = shift @poly_fasta;
$dsim = shift @poly_fasta;

#print $poly_fasta[0];

#Retirar la secuencia del outgroup
#if ($poly_fasta[0] =~/^.*FBtr\d+.*_dsim.+/){
# $dsim = shift @poly_fasta;
#}

#Convertir el array en variable para poder pasarlo por la subrutina. Cada linea id\tseq
my $fh = join(' ', @poly_fasta);
$fh =~s/\t\t/\n/g;
$fh =~s/\t/\n/g;
#print RESULTS $fh;

#Obtener la secuencia consenso del resto de las secuencias, las polimórficas
#my $fasta = get_next_record($fh);
my $raw = fasta2raw($fh);
#print $raw;
my $self = Bio::UnivAln->new(-seqs=>$raw);
my $consensus = greedy($self, 0.5, 1);

# abrimos el fichero donde vamos a escribir las secuencias consenso en fasta
$dere =~s/\t/\n/g;
$recoded =~s/\t/\n/g;
#print RESULTS ">".$recoded;
print RESULTS ">".$dere;
print RESULTS ">poly_cons_seq"."\\n".$consensus."\\n";

#print 'The process has finished.';
close RESULTS;

exit;

```

## ARTICLE

doi:10.1038/nature10811

# The *Drosophila melanogaster* Genetic Reference Panel

Trudy F. C. Mackay<sup>1\*</sup>, Stephen Richards<sup>2\*</sup>, Eric A. Stone<sup>3\*</sup>, Antonio Barbadilla<sup>3\*</sup>, Julien F. Ayroles<sup>1†</sup>, Dianhui Zhu<sup>2</sup>, Sónia Casillas<sup>3†</sup>, Yi Han<sup>2</sup>, Michael M. Magwire<sup>1</sup>, Julie M. Cridland<sup>4</sup>, Mark F. Richardson<sup>5</sup>, Robert R. H. Anholt<sup>6</sup>, Maite Barrón<sup>2</sup>, Crystal Bess<sup>2</sup>, Kerstin Petra Blankenburg<sup>2</sup>, Mary Anna Carbone<sup>1</sup>, David Castellano<sup>3</sup>, Lesley Chaboub<sup>2</sup>, Laura Duncan<sup>1</sup>, Zeke Harris<sup>1</sup>, Mehwish Javald<sup>2</sup>, Joy Christina Jayaseelan<sup>2</sup>, Shalini N. Jhangiani<sup>2</sup>, Katherine W. Jordan<sup>1</sup>, Fremiet Lara<sup>2</sup>, Faye Lawrence<sup>1</sup>, Sandra L. Lee<sup>2</sup>, Pablo Librado<sup>7</sup>, Raquel S. Linheiro<sup>3</sup>, Richard F. Lyman<sup>1</sup>, Aaron J. Mackey<sup>8</sup>, Mala Munidasa<sup>2</sup>, Donna Marie Muzny<sup>2</sup>, Lynne Nazareth<sup>2</sup>, Irene Newsham<sup>2</sup>, Lora Perales<sup>2</sup>, Ling-Ling Pu<sup>2</sup>, Carson Qu<sup>2</sup>, Miquel Ràmia<sup>3</sup>, Jeffrey G. Reid<sup>2</sup>, Stephanie M. Rollmann<sup>1†</sup>, Julio Rozas<sup>2</sup>, Nehad Saada<sup>2</sup>, Lavanya Turlapati<sup>1</sup>, Kim C. Worley<sup>2</sup>, Yuan-Qing Wu<sup>2</sup>, Akihiko Yamamoto<sup>1</sup>, Yiming Zhu<sup>2</sup>, Casey M. Bergman<sup>9</sup>, Kevin R. Thornton<sup>4</sup>, David Mittelman<sup>9</sup> & Richard A. Gibbs<sup>2</sup>

A major challenge of biology is understanding the relationship between molecular genetic variation and variation in quantitative traits, including fitness. This relationship determines our ability to predict phenotypes from genotypes and to understand how evolutionary forces shape variation within and between species. Previous efforts to dissect the genotype-phenotype map were based on incomplete genotypic information. Here, we describe the *Drosophila melanogaster* Genetic Reference Panel (DGRP), a community resource for analysis of population genomics and quantitative traits. The DGRP consists of fully sequenced inbred lines derived from a natural population. Population genomic analyses reveal reduced polymorphism in centromeric autosomal regions and the X chromosome, evidence for positive and negative selection, and rapid evolution of the X chromosome. Many variants in novel genes, most at low frequency, are associated with quantitative traits and explain a large fraction of the phenotypic variance. The DGRP facilitates genotype-phenotype mapping using the power of *Drosophila* genetics.

Understanding how molecular variation maps to phenotypic variation for quantitative traits is central for understanding evolution, animal and plant breeding, and personalized medicine<sup>1,2</sup>. The principles of mapping quantitative trait loci (QTLs) by linkage to, or association with, marker loci are conceptually simple<sup>1,2</sup>. However, we have not yet achieved our goal of explaining genetic variation for quantitative traits in terms of the underlying genes; additive, epistatic and pleiotropic effects as well as phenotypic plasticity of segregating alleles; and the molecular nature, population frequency and evolutionary dynamics of causal variants. Efforts to dissect the genotype-phenotype map in model organisms<sup>3,4</sup> and humans<sup>5-7</sup> have revealed unexpected complexities, implicating many, novel loci, pervasive pleiotropy, and context-dependent effects.

Model organism reference populations of inbred strains that can be shared among laboratories studying diverse phenotypes, and for which environmental conditions can be controlled and manipulated, greatly facilitate efforts to dissect the genetic architecture of quantitative traits<sup>3,4</sup>. Measuring many individuals of the same homozygous genotype increases the accuracy of the estimates of genotypic value<sup>1</sup> and the power to detect variants, and genotypes of molecular markers need only be obtained once. We constructed the *Drosophila melanogaster* Genetic Reference Panel (DGRP) as such a community resource. Unlike previous populations of recombinant inbred lines derived from limited samples of genetic variation, the DGRP consists

of 192 inbred strains derived from a single outbred population. The DGRP contains a representative sample of naturally segregating genetic variation, has an ultra-fine-grained recombination map suitable for precise localization of causal variants, and has almost complete euchromatic sequence information.

Here, we describe molecular and phenotypic variation in 168 resequenced lines comprising Freeze 1.0 of the DGRP, population genomic inferences of patterns of polymorphism and divergence and their correlation with genomic features, local recombination rate and selection acting on this population, genome-wide association mapping analyses for three quantitative traits, and tools facilitating the use of this resource.

## Molecular variation in the DGRP

We constructed the DGRP by collecting mated females from the Raleigh, North Carolina, USA, population, followed by 20 generations of full-sibling inbreeding of their progeny. We sequenced 168 DGRP lines using a combination of Illumina and 454 sequencing technology: 29 of the lines were sequenced using both platforms, 129 lines have only Illumina sequence, and 10 lines have only 454 sequence. We mapped sequence reads to the *D. melanogaster* reference genome, re-calibrated base quality scores, and locally re-aligned Illumina reads. Mean sequence coverage was 21.4× per line for Illumina sequences and 12.1× per line for 454 sequences (Supplementary

<sup>1</sup>Department of Genetics, North Carolina State University, Raleigh, North Carolina 27695, USA. <sup>2</sup>Human Genome Sequencing Center, Baylor College of Medicine, One Baylor Plaza, Houston, Texas 77030 USA. <sup>3</sup>Genomics, Bioinformatics and Evolution Group, Institut de Biociències i de Biomedicina - IBB/Department of Genetics and Microbiology, Campus Universitat Autònoma de Barcelona, 08193 Bellaterra, Spain. <sup>4</sup>Department of Ecology and Evolutionary Biology, University of California - Irvine, Irvine, California 92697, USA. <sup>5</sup>Faculty of Life Sciences, University of Manchester, Manchester M13 9PT, UK. <sup>6</sup>Department of Biology, North Carolina State University, Raleigh, North Carolina 27695, USA. <sup>7</sup>Molecular Evolutionary Genetics Group, Department of Genetics, Faculty of Biology, Universitat de Barcelona, Diagonal 645, 08028 Barcelona, Spain. <sup>8</sup>Center for Public Health Genomics, University of Virginia, PO Box 800717, Charlottesville, Virginia 22908, USA. <sup>9</sup>Virginia Bioinformatics Institute and Department of Biological Sciences, Virginia Tech, Blacksburg, Virginia 24061, USA. <sup>†</sup>Present addresses: FAS Society of Fellows, Harvard University, 78 Mt Auburn Street, Cambridge, Massachusetts 02138, USA (J.F.A.); Functional Comparative Genomics Group, Institut de Biociències i de Biomedicina - IBB, Campus Universitat Autònoma de Barcelona, 08193 Bellaterra, Spain (S.C.); Department of Biological Sciences, University of Cincinnati, Cincinnati, Ohio 45221, USA (S.M.R.).

\*These authors contributed equally to this work.

Table 1). On average, we assayed 113.5 megabases (94.25%) of the euchromatic reference sequence with ~22,000 read mapping gaps per line (Supplementary Table 2). We called 4,672,297 single nucleotide polymorphisms (SNPs) using the Joint Genotyper for Inbred Lines (JGIL; E.A.S., personal communication), which takes into account coverage and quality sequencing statistics, and expected allele frequencies after 20 generations of inbreeding from an outbred population initially in Hardy–Weinberg equilibrium. In cases where base calls were made by both technologies, concordance was 99.36% (Supplementary Table 3).

The SNP site frequency distribution (Fig. 1a) is characterized by a majority of low frequency variants. The numbers of SNPs vary by chromosome and site class (Fig. 1b). Linkage disequilibrium<sup>8</sup> decays to  $r^2 = 0.2$  on average within 10 base pairs on autosomes and 30 base pairs on the X chromosome (Fig. 1c and Supplementary Fig. 1). This difference is expected because the population size of the X chromosome is three quarters that of autosomes, and the X chromosome can experience greater purifying selection because of exposure of deleterious recessive alleles in hemizygous males. There is little evidence of global population structure in the DGRP (Fig. 1d and Supplementary Fig. 2). The rapid decline in linkage disequilibrium locally and lack of global population structure are favourable for genome-wide association mapping.

Not all SNPs are fixed within individual DGRP lines (Supplementary Table 4). The expected inbreeding coefficient ( $F$ ) after 20 generations of full-sib inbreeding<sup>1</sup> is  $F = 0.986$ ; therefore, we expect some SNPs to remain segregating by chance. Segregating SNPs can also arise from new mutations, or if natural selection opposes inbreeding, due to true overdominance for fitness at individual loci or associative overdominance due to complementary deleterious alleles that are closely linked or in segregating inversions.

We identified 390,873 microsatellite loci, 105,799 of which were polymorphic (Supplementary Table 5); 36,810 transposable element insertion sites and 197,402 total insertions (Supplementary Table 6). On average, each line contained 1,175 transposable element insertions (Supplementary Table 6), although most transposable element insertion sites (25,562) were present in only one line (Supplementary

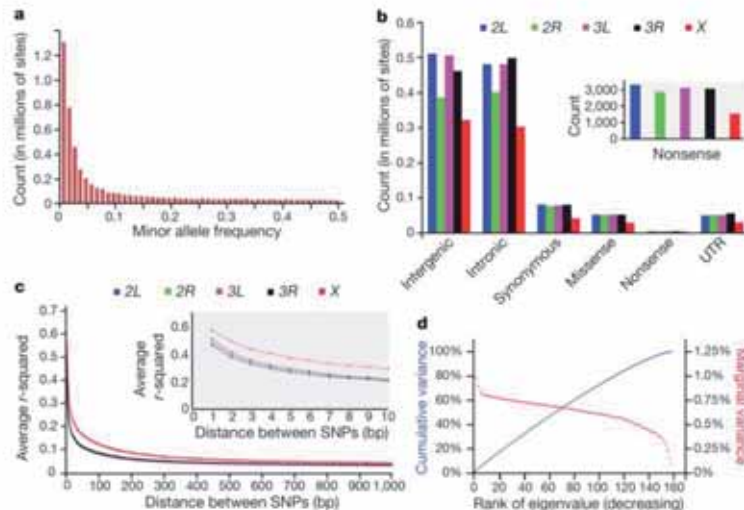
Table 7). We identified 149 transposable element families. The number of copies per family varied greatly from an average of 315.7 *INE-1* elements per line to an average of 0.003 *Gandalf-Dkoe-like* elements per line (Supplementary Table 8).

*Wolbachia pipiensis* is a maternally inherited bacterium found in insects, including *Drosophila*, and can affect reproduction<sup>9</sup>. We assessed *Wolbachia* infection status in the DGRP lines to account for it in analyses of genotype–phenotype associations, and found 51.2% of lines harbouring sufficient *Wolbachia* DNA to imply infection (Supplementary Table 9).

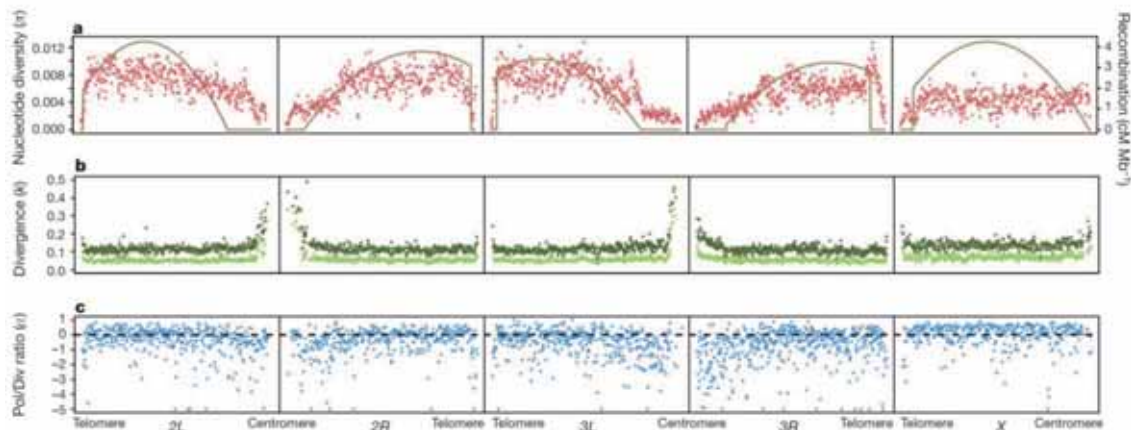
### Polymorphism and divergence

We used the DGRP Illumina sequence data and genome sequences from *Drosophila simulans* and *Drosophila yakuba*<sup>10</sup> to perform genome-wide analyses of polymorphism and divergence, assess the association of these parameters with genomic features and the recombination landscape, and infer the historical action of selection on a much larger scale than had been possible previously<sup>11–16</sup>. We computed polymorphism ( $\pi$  and  $\theta$ , refs 17 and 18) and divergence ( $k$ , ref. 19) for the whole genome, by chromosome arm (X, 2L, 2R, 3L, 3R), by chromosome region (three regions of equal size in Mb—telomeric, middle and centromeric), in 50-kbp non-overlapping windows, and by site class (synonymous and non-synonymous sites within coding sequences, and intronic, untranslated region (UTR) and intergenic sites) (Supplementary Tables 10 and 11).

Averaged over the entire genome,  $\pi = 0.0056$  and  $\theta = 0.0067$ , similar to previous estimates from North American populations<sup>16,20</sup>. Average polymorphism on the X chromosome ( $\pi_X = 0.0040$ ) is reduced relative to the autosomes ( $\pi_A = 0.0060$ ) (X/A ratio = 0.67, Wilcoxon test  $P = 0$ ), even after correcting for the X/A effective population size ( $X_{4/3} = 0.0054$ , Wilcoxon test  $P < 0.00002$ ; Supplementary Table 10). Autosomal nucleotide diversity is reduced on average 2.4-fold in centromeric regions relative to non-centromeric regions, and at the telomeres (Fig. 2a and Supplementary Table 10), whereas diversity is relatively constant along the X chromosome. Thus,  $\pi_X > \pi_A$  in centromeric regions, but  $\pi_A > \pi_X$  in other chromosomal regions (Fig. 2a and Supplementary Table 10).



**Figure 1** | SNP variation in the DGRP lines. **a**, Site frequency spectrum. **b**, Numbers of SNPs per site class. **c**, Decay of linkage disequilibrium ( $r^2$ ) with physical distance for the five major chromosome arms. **d**, Lack of population structure. The red curve depicts the ranked eigenvalues of the genetic covariance matrix in decreasing order with respect to the marginal variance explained; the blue curve shows their cumulative sum as a fraction of the total with respect to cumulative variance explained. The partitioning of total genetic variance is balanced among the eigenvectors. The principal eigenvector explains < 1.1% of the total genetic variance.



**Figure 2 | Pattern of polymorphism, divergence,  $\alpha$  and recombination rate along chromosome arms in non-overlapping 50-kbp windows.** **a**, Nucleotide polymorphism ( $\pi$ ). The solid curves give the recombination rate ( $\text{cM Mb}^{-1}$ ). **b**, Divergence ( $k$ ) for *D. simulans* (light green) and *D. yakuba* (dark green). **c**, Polymorphism to divergence ratio (Pol/Div), estimated as  $1 - [(\pi_{0-0.001}/\pi_{4-0.001}) / (k_{0-0.001}/k_{4-0.001})]$ . An excess of 0-fold divergence relative to polymorphism ( $k_{0-0.001}/k_{4-0.001} > (\pi_{0-0.001}/\pi_{4-0.001})$ ) is interpreted as adaptive fixation whereas an excess of 0-fold polymorphism relative to divergence ( $\pi_{0-0.001}/\pi_{4-0.001} > (k_{0-0.001}/k_{4-0.001})$ ) indicates that weakly deleterious or nearly neutral mutations are segregating in the population.

Genes on the X chromosome evolve faster ( $k_X = 0.140$ ) than autosomal genes ( $k_A = 0.126$ ) ( $X/A$  ratio = 1.131, Wilcoxon test  $P = 0$ ) (Fig. 2b and Supplementary Table 10). Divergence is more uniform (coefficient of variation  $(CV)_k = 0.2841$ ) across chromosome arms than is polymorphism ( $(CV)_\pi = 0.4265$ ). The peaks of divergence near the centromeres could be attributable to the reduced quality of alignments in these regions. Patterns of divergence are similar regardless of the outgroup species used (Fig. 2b and Supplementary Table 11).

The pattern of polymorphism and divergence by site class is consistent within and among chromosomes ( $\pi_{k_{\text{synonymous}}} > \pi_{k_{\text{missense}}} > \pi_{k_{\text{nonsynonymous}}} > \pi_{k_{\text{stop}}}$ ), in agreement with previous studies on smaller data sets<sup>12,13</sup> (Supplementary Figs 3 and 4 and Supplementary Table 11). Polymorphism levels between synonymous and non-synonymous sites differ by an order of magnitude. Variation and divergence patterns within the site classes generally follow the same patterns observed overall, with reduced polymorphism for all site classes on the X chromosome relative to autosomes, increased X chromosome divergence relative to autosomes for all but synonymous sites, decreased polymorphism in centromeric regions, and greater variation among regions and arms for polymorphism than for divergence. Other diversity measures and more detailed patterns at different window-sizes for each chromosome arm can be accessed from the Population *Drosophila* Browser (popDrowser) (Table 1 and Methods).

### Recombination landscape

Evolutionary models of hitchhiking and background selection<sup>21,22</sup> predict a positive correlation between polymorphism and recombination rate. This expectation is realized in regions where recombination is less than  $2 \text{ cM Mb}^{-1}$  (Spearman's  $\rho = 0.471$ ,  $P = 0$ ), but recombination exceeds  $2 \text{ cM Mb}^{-1}$  (Spearman's  $\rho = -0.0044$ ,  $P = 0.987$ ) (Fig. 2a and Supplementary Table 12). The average rate of recombination of the X chromosome ( $2.9 \text{ cM Mb}^{-1}$ ) is greater than that of autosomes ( $2.1 \text{ cM Mb}^{-1}$ ), which may account for the low overall X-linked correlation between recombination rate and  $\pi$ . The lack of correlation between recombination and divergence (Supplementary Table 12) excludes mutation associated with recombination as the cause of the correlation. We assessed the independent effects of recombination rate, divergence, chromosome region and gene density on nucleotide variation of autosomes and the X chromosome (Supplementary Table 13). Recombination is the major predictor of

polymorphism on the X chromosome and autosomes; however, the significant effect of autosomal chromosome region remains after accounting for variation in recombination rates between centromeric and non-centromeric regions.

### Selection regimes

We used the standard<sup>23</sup> and generalized<sup>12,24,25</sup> McDonald Kreitman tests (MKT) to scan the genome for evidence of selection. These tests

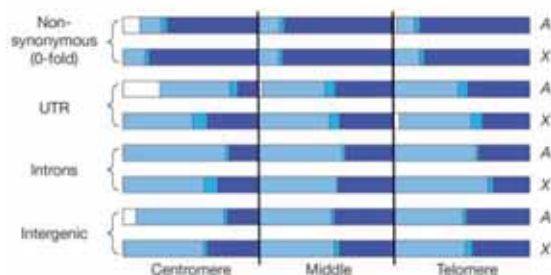
**Table 1 | Community resources**

Resource	Location
DGRP lines	Bloomington <i>Drosophila</i> Stock Center <a href="http://flystocks.bio.indiana.edu/Browse/RAL.php">http://flystocks.bio.indiana.edu/Browse/RAL.php</a>
Sequences	Baylor College of Medicine Human Genome Sequencing Center <a href="http://www.hgsc.bcm.tmc.edu/project-species-i-DGRP_lines.hgsc">http://www.hgsc.bcm.tmc.edu/project-species-i-DGRP_lines.hgsc</a> National Center for Biotechnology Information Short Read Archive <a href="http://www.ncbi.nlm.nih.gov/sra/?term=DGRP">http://www.ncbi.nlm.nih.gov/sra/?term=DGRP</a> Mackay Laboratory <a href="http://dgrp.gnets.ncsu.edu/">http://dgrp.gnets.ncsu.edu/</a>
Read alignments	Baylor College of Medicine Human Genome Sequencing Center <a href="http://www.hgsc.bcm.tmc.edu/projects/dgrp/">http://www.hgsc.bcm.tmc.edu/projects/dgrp/</a> Baylor College of Medicine Human Genome Sequencing Center <a href="http://www.hgsc.bcm.tmc.edu/projects/dgrp/freeze1_July_2010/snp_calls/">http://www.hgsc.bcm.tmc.edu/projects/dgrp/freeze1_July_2010/snp_calls/</a> National Center for Biotechnology Information dbSNP <a href="http://www.ncbi.nlm.nih.gov/SNP/snp_viewBatch.cgi?sbid=1052186">http://www.ncbi.nlm.nih.gov/SNP/snp_viewBatch.cgi?sbid=1052186</a> Mackay Laboratory <a href="http://dgrp.gnets.ncsu.edu/">http://dgrp.gnets.ncsu.edu/</a>
SNPs	Baylor College of Medicine Human Genome Sequencing Center <a href="http://www.hgsc.bcm.tmc.edu/projects/dgrp/freeze1_July_2010/snp_calls/">http://www.hgsc.bcm.tmc.edu/projects/dgrp/freeze1_July_2010/snp_calls/</a> National Center for Biotechnology Information dbSNP <a href="http://www.ncbi.nlm.nih.gov/SNP/snp_viewBatch.cgi?sbid=1052186">http://www.ncbi.nlm.nih.gov/SNP/snp_viewBatch.cgi?sbid=1052186</a> Mackay Laboratory <a href="http://dgrp.gnets.ncsu.edu/">http://dgrp.gnets.ncsu.edu/</a>
Microsatellites	Baylor College of Medicine Human Genome Sequencing Center <a href="http://www.hgsc.bcm.tmc.edu/projects/dgrp/freeze1_July_2010/microsat_calls/">http://www.hgsc.bcm.tmc.edu/projects/dgrp/freeze1_July_2010/microsat_calls/</a> Mittelman Laboratory <a href="http://genome.vbi.vt.edu/public/DGRP/">http://genome.vbi.vt.edu/public/DGRP/</a>
Transposable elements	Mackay Laboratory <a href="http://dgrp.gnets.ncsu.edu/">http://dgrp.gnets.ncsu.edu/</a>
Molecular population genomics	PopDrowser <a href="http://popdrowser.uab.cat">http://popdrowser.uab.cat</a>
Phenotypes	Mackay Laboratory <a href="http://dgrp.gnets.ncsu.edu/">http://dgrp.gnets.ncsu.edu/</a>
Genome-wide association analysis	Mackay Laboratory <a href="http://dgrp.gnets.ncsu.edu/">http://dgrp.gnets.ncsu.edu/</a>

compare the ratio of polymorphism at a selected site with that of a neutral site to the ratio of divergence at a selected site to divergence at a neutral site. The standard MKT is applied to coding sequences, and synonymous and non-synonymous sites are used as putative neutral and selected sites, respectively. The generalized MKT is applied to non-coding sequences and uses fourfold degenerate sites as neutral sites. Using polymorphism and divergence data avoids confounding inference of selection with mutation rate differences, and restricting the tests to closely linked sites controls for shared evolutionary history<sup>28–30</sup>. We infer adaptive divergence when there is an excess of divergence relative to polymorphism, and segregation of slightly deleterious mutations when there is an excess of polymorphism over divergence. Estimates of  $\alpha$ , the proportion of adaptive divergence, are biased downwards by low frequency, slightly deleterious mutations<sup>28,30</sup>. Rather than eliminate low frequency variants<sup>31</sup>, we incorporated information on the site frequency distribution to the MKT test framework to obtain estimates of the proportion of sites that are strongly deleterious ( $d$ ), weakly deleterious ( $b$ ), neutral ( $f$ ) and recently neutral ( $\gamma$ ) at segregating sites, as well as unbiased estimates of  $\alpha$  (Supplementary Methods).

### Deleterious and neutral sites

Averaged over the entire genome, we infer that 58.5% of the segregating sites are neutral or nearly neutral, 1.9% are weakly deleterious and 39.6% are strongly deleterious. However, these proportions vary between the X chromosome and autosomes, site classes and chromosome regions (Supplementary Tables 14–16 and Fig. 3). Non-synonymous sites are the most constrained ( $d = 77.6\%$ ), whereas in non-coding sites  $d$  ranges from 29.1% in 5' UTRs to 41.3% in 3' intergenic regions. The inferred pattern of selection differs between autosomal centromeric and non-centromeric regions:  $d$  is reduced and  $f$  is increased in centromeric regions for all site categories (Fig. 3). We observe an excess of polymorphism relative to divergence in autosomal centromeric regions, even after correcting for weakly deleterious mutations, implying a relaxation of selection from the time of separation of *D. melanogaster* and *D. yakuba*. Because selection coefficients depend on the effective population size<sup>32</sup> ( $N_e$ ), this could occur if the recombination rate has specifically diminished in centromeric regions during the divergence between *D. melanogaster* and *D. yakuba*; or with an overall reduction of  $N_e$  associated with the colonization of North American habitats<sup>33,34</sup>. In the latter case, we expect a genome-wide signature of an excess of low-frequency polymorphisms and of polymorphism relative to divergence, exacerbated in regions of low recombination. We indeed find an excess of low-frequency polymorphism relative to neutral expectation as indicated by the negative estimates of Tajima's  $D$  statistic<sup>35</sup>



**Figure 3** | The fraction of alleles segregating under different selection regimes by site class and chromosome region, for the autosomes (A) and the X chromosome (X). The selection regimes are strongly deleterious ( $d$ , dark blue), weakly deleterious ( $b$ , blue), recently neutral ( $\gamma$ , white) and old neutral ( $f$ , light blue). Each chromosome arm has been divided in three regions of equal size (in Mb): centromere, middle and telomere.

( $D = -0.686$  averaged over the whole genome and  $D = -0.997$  in autosomal centromeric regions). In contrast, the X chromosome does not show a differential pattern of selection in the centromeric region, has a lower fraction of relaxation of selection, fewer neutral alleles, and a higher percentage of strongly deleterious alleles for all site classes and regions (Fig. 3 and Supplementary Tables 14–16).

Transposable element insertions are thought to be largely deleterious. There are more singleton insertions in regions of high recombination ( $\geq 2 \text{ cM Mb}^{-1}$ ) and more insertions shared in multiple lines in regions of low recombination ( $< 2 \text{ cM Mb}^{-1}$ ) (Fisher's exact test  $P = 0$ ), and comparison of observed and expected site occupancy spectra reveals an excess of singleton insertions ( $P = 0$ , Supplementary Fig. 5).

### Adaptive fixation

We find substantial evidence for positive selection in autosomal non-centromeric regions and the X chromosome (Fig. 2c and Supplementary Tables 15 and 17). We estimated  $\alpha$  by aggregating all sites in each region analysed to avoid underestimation by averaging across genes<sup>36</sup> in comparisons of chromosomes, regions and site classes. We also computed the direction of selection, DoS<sup>37</sup>, which is positive with adaptive selection, zero under neutrality and negative when weakly deleterious or new nearly neutral mutations are segregating. Estimates of  $\alpha$  from the standard and generalized MKT indicate that on average 25.2% of the fixed sites between *D. melanogaster* and *D. yakuba* are adaptive, ranging from 30% in introns to 7% in UTR sites (Supplementary Fig. 6). Estimates of DoS and  $\alpha$  are negative for non-synonymous and UTR sites in the autosomal centromeres, consistent with underestimating the fraction of adaptive substitutions in regions of low recombination because weakly deleterious or nearly neutral mutations are more common than adaptive fixations. The majority of adaptive fixation on autosomes occurs in non-centromeric regions (Fig. 2c). We find over four times as many adaptive fixations on the X chromosome relative to autosomes. The pattern holds for all site classes, in particular non-synonymous sites and UTRs, as well as individual genes, and is not solely due to the autosomal centromeric effect (Supplementary Table 15 and Supplementary Figs 6 and 7). Finally, when we consider DoS in recombination environments above and below  $2 \text{ cM Mb}^{-1}$ , we find greater adaptive propensity in genes whose recombination context is  $\geq 2 \text{ cM Mb}^{-1}$  (Wilcoxon test,  $P = 0$ ; Supplementary Fig. 8).

To understand the global patterns of divergence and constraint across functional classes of genes, we examined the distributions of  $\omega$  ( $d_n/d_s$ , the ratio of non-synonymous to synonymous divergence) and DoS across gene ontology (GO) categories. The 4.9% GO categories with significantly elevated DoS include the biological process categories of behaviour, developmental process involved in reproduction, reproduction and ion transport (Supplementary Table 18). Recombination context is the major determinant of variation in DoS (Supplementary Table 19) whereas GO category is as important as recombinational context for predicting variation in  $\omega$  (Supplementary Table 19).

GO categories enriched for positive DoS values differ from those associated with high values of  $\omega$  (Supplementary Table 18), indicating that positive selection does not occur necessarily on genes with high  $\omega$  values. If adaptive substitutions are common, high values of  $\omega$  reflect the joint contributions of neutral and adaptive substitutions. Further, equating high constraint (low  $\omega$ ) with functional importance overlooks the functional role of adaptive changes<sup>35</sup>. Unlike  $\omega$ , DoS takes into account the constraints inferred from the current polymorphism, distinguishing negative, neutral and adaptive selection.

### Genome-wide genotype–phenotype associations

We measured resistance to starvation stress, chill coma recovery time and startle response<sup>38</sup> in the DGRP. We found considerable genetic variation for all traits, with high broad sense heritabilities. We also found variation in sex dimorphism for starvation resistance and chill



coma recovery with cross-sex genetic correlations significantly different from unity (Supplementary Tables 20–22).

We performed genome-wide association analyses for these traits, using the 2,490,165 SNPs and 77,756 microsatellites for which the minor allele was represented in four or more lines, using single-locus analyses pooled across sexes and separately for males and females. At  $P < 10^{-5}$  ( $P < 10^{-6}$ ), we find 203 (32) SNPs and 2 (0) microsatellites associated with starvation resistance; 90 (7) SNPs and 4 (2) microsatellites associated with startle response; and 235 (45) SNPs and 5 (3) microsatellites associated with chill coma recovery time (Fig. 4a, Supplementary Fig. 9 and Supplementary Tables 23 and 24). The minor allele frequencies for most of the associated SNPs are low, and there is an inverse relationship between effect sizes and minor allele frequency (Supplementary Fig. 10).

The DGRP is a powerful tool for rapidly reducing the search space for molecular variants affecting quantitative traits from the entire genome to candidate polymorphisms and genes. Although we cannot infer which of these polymorphisms are causal due to linkage disequilibrium between SNPs in close physical proximity as well as occasional spurious long range linkage disequilibrium (Fig. 4a and Supplementary Fig. 9), the candidate gene lists are likely to be enriched for causal variants. The majority of associations are in computationally predicted genes or genes with annotated functions not obviously associated with the three traits. However, genes previously associated with startle response<sup>20</sup> (*Sema-1a* and *Eip75B*) and starvation resistance<sup>21</sup> (*pnt*) were identified in this study; and a SNP in *CG3213*, previously identified in a *Drosophila* obesity screen<sup>22</sup>, is associated with variation in starvation resistance. Several genes associated with quantitative traits are rapidly evolving (*psg*, *Egfr*; Supplementary Tables 17 and 23) or are plausible candidates based on SNP or gene ontology annotations (Supplementary Table 23).

### Predicting phenotypes from genotypes

We used regression models to predict trait phenotypes from SNP genotypes and estimate the total variance explained by SNPs. The latter cannot be done by summing the individual contributions of the single marker effects because markers are not completely independent, and estimates of effects of single markers are biased when more than one locus affecting the trait segregates in the population. We derived gene-centred multiple regression models to estimate the effects of multiple SNPs simultaneously. In all cases 6–10 SNPs explain from 51–72% of the phenotypic variance and 65–90% of the genetic variance (Supplementary Tables 25 and 26 and Supplementary Figs 11–13). We also derived partial least square regression models using all SNPs for which the single marker effect was significant

at  $P < 10^{-5}$ . These models explain 72–85% of the phenotypic variance (Fig. 4b, c and Supplementary Fig. 14).

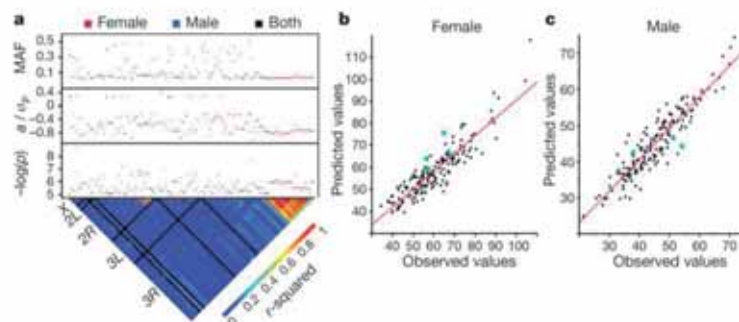
### Discussion

The DGRP lines, sequences, variant calls, phenotypes and web tools for molecular population genomics and genome-wide association analysis are publicly available (Table 1). The DGRP lines contain at least 4,672,297 SNPs, 105,799 polymorphic microsatellites and 36,810 transposable elements, as well as insertion/deletion events and copy number variants and are a valuable resource for understanding the genetic architecture of quantitative traits of ecological and evolutionary relevance as well as *Drosophila* models of human quantitative traits. These novel mutations have survived the sieve of natural selection and will enhance the functional annotation of the *Drosophila* genome, complementing the *Drosophila* Gene Disruption Project<sup>23</sup> and the *Drosophila* modENCODE project<sup>24</sup>.

Genome-wide molecular population genetic analyses show that patterns of polymorphism, but not divergence, differ by autosomal chromosome region, and between the X chromosome and autosomes. Polymorphism is lower in autosomal centromeric than non-centromeric regions, but not for the X chromosome. We propose that the correlation of polymorphism with recombination in regions where recombination is  $< 2 \text{ cM Mb}^{-1}$  is due to the reduced effective population size in regions of low recombination<sup>25</sup>. Selection is less efficient in regions of low recombination<sup>26</sup>, consistent with our observation that the fraction of strongly deleterious mutations and positively selected sites are reduced in these regions.

All molecular population genomic analyses support the 'faster X' hypothesis<sup>27</sup>. Relative to the autosomes, the X chromosome shows lower polymorphism, faster rates of molecular evolution, a higher percentage of gene regions undergoing adaptive evolution, a higher fraction of strongly deleterious sites, and a lower level of weak negative selection and relaxation of selection. New X-linked mutations are directly exposed to selection each generation in hemizygous males, and the X chromosome has greater recombination than autosomes<sup>28</sup>; both of these factors could contribute to this observation.

Genome-wide association analyses of three fitness-related quantitative traits reveal hundreds of novel candidate genes, highlighting our ignorance of the genetic basis of complex traits. Most variants associated with the traits are at low frequency, and there is an inverse relationship between frequency and effect. Given that low-frequency alleles are likely to be deleterious for traits under directional or stabilizing selection, these results are consistent with the mutation-selection balance hypothesis<sup>29</sup> for the maintenance of quantitative genetic variation. Regression models incorporating significant SNPs



**Figure 4 | Genotype-phenotype associations for starvation resistance.** **a**, Genome-wide association results for significant SNPs. The lower triangle depicts linkage disequilibrium ( $r^2$ ) among SNPs, with the five major chromosome arms demarcated by black lines. The upper panels give the significance threshold ( $-\log(p)$ , uncorrected for multiple tests), the effect in phenotypic standard deviation units, and the minor allele frequency (MAF). **b**, **c**, Partial least squares regressions of phenotypes predicted using SNP data on observed phenotypes. The blue dots represent the predicted and observed phenotypes of lines that were not included in the initial study. **b**, Females ( $r^2 = 0.81$ ); **c**, males ( $r^2 = 0.85$ ).

explain most of the phenotypic variance of the traits, in contrast with human association studies, where significant SNPs have tiny effects and together explain a small fraction of the total phenotypic variance<sup>7</sup>. If the genetic architecture of human complex traits is also dominated by low-frequency causal alleles, we expect estimates of effect size based on linkage disequilibrium with common variants to be strongly biased downwards.

In the future, the full power of *Drosophila* genetics can be applied to validating marker-trait associations: mutations, RNA interference constructs and quantitative trait loci mapping populations. The DGRP is an ideal resource for systems genetics analyses of the relationship between molecular variation, causal molecular networks and genetic variation for complex traits<sup>4,38,49</sup>, and will anchor evolutionary studies in comparison with sequenced *Drosophila* species to assess to what extent variation within a species corresponds to variation among species.

## METHODS SUMMARY

The full Methods are in the Supplementary Information. Information on sequencing and bioinformatics includes methods for DNA isolation; library construction and genomic sequencing; sequence read alignment; SNP, microsatellite and transposable element identification; genotypes for assurance of sample identity; and *Wolbachia* detection. Methods for molecular population genomics analysis include details of recombination estimates; diversity measures, linkage disequilibrium and neutrality tests; software used for population genomic analysis; data visualization (popDrowser); standard and generalized McDonald-Kreitman tests, statistical analysis methods; quality assessment and data filtering; and gene ontology analyses. Methods for quantitative genetic analyses include phenotype measures, quantitative genetic analyses of phenotypes, statistical analyses of genotype-phenotype associations and predictive models, and a web-based association analysis pipeline.

Received 13 July; accepted 21 December 2011.

- Falconer, D. S. & Mackay, T. F. C. *Introduction to Quantitative Genetics* 4th edn (Longman, 1996).
- Lynch, M. & Walsh, B. *Genetics and Analysis of Quantitative Traits* (Sinauer Associates, 1998).
- Flint, J. & Mackay, T. F. C. Genetic architecture of quantitative traits in flies, mice and humans. *Genome Res.* **19**, 723–733 (2009).
- Mackay, T. F. C., Stone, E. A. & Ayroles, J. F. The genetics of quantitative traits: challenges and prospects. *Nature Rev. Genet.* **10**, 565–577 (2009).
- Altshuler, D., Daly, M. J. & Lander, E. S. Genetic mapping in human disease. *Science* **322**, 881–888 (2008).
- Donnelly, P. Progress and challenges in genome-wide association studies in humans. *Nature* **456**, 728–731 (2008).
- Manolio, T. A. et al. Finding the missing heritability of complex diseases. *Nature* **461**, 747–753 (2009).
- Hill, W. G. & Robertson, A. The effect of linkage on limits to artificial selection. *Genet. Res.* **8**, 269–294 (1966).
- Warren, J. H. Biology of *Wolbachia*. *Annu. Rev. Entomol.* **42**, 587–609 (1997).
- Clark, A. G. et al. Evolution of genes and genomes on the *Drosophila* phylogeny. *Nature* **450**, 203–218 (2007).
- Smith, N. G. & Eyre-Walker, A. Adaptive protein evolution in *Drosophila*. *Nature* **415**, 1022–1024 (2002).
- Andolfatto, P. Adaptive evolution of non-coding DNA in *Drosophila*. *Nature* **437**, 1149–1152 (2005).
- Presgraves, D. C. Recombination enhances protein adaptation in *Drosophila melanogaster*. *Curr. Biol.* **15**, 1651–1656 (2005).
- Casillas, S., Barbadilla, A. & Bergman, C. Purifying selection maintains highly conserved noncoding sequences in *Drosophila*. *Mol. Biol. Evol.* **24**, 2222–2234 (2007).
- Sella, G. et al. Pervasive natural selection in the *Drosophila* genome? *PLoS Genet.* **5**, e1000495 (2009).
- Sackton, T. B. et al. Population genomic inferences from sparse high-throughput sequencing of two populations of *Drosophila melanogaster*. *Genome Biol. Evol.* **1**, 449–465 (2009).
- Nei, M. *Molecular Evolutionary Genetics* (Columbia Univ. Press, 1987).
- Watterson, G. A. On the number of segregating sites in genetical models without recombination. *Theor. Popul. Biol.* **7**, 256–276 (1975).
- Jukes, T. H. & Cantor, C. R. In *Mammalian Protein Metabolism* vol. 3 (eds Munro, H. N. & Allison, J. B.) 21–132 (Academic Press, 1969).
- Andolfatto, P. & Przeworski, M. Regions of lower crossing over harbor more rare variants in African *Drosophila melanogaster*. *Genetics* **158**, 657–665 (2001).

- Begun, D. J. & Aquadro, C. F. Levels of naturally occurring DNA polymorphism correlate with recombination rates in *D. melanogaster*. *Nature* **356**, 519–520 (1992).
- Charlesworth, B., Morgan, M. T. & Charlesworth, D. The effect of deleterious mutations on neutral molecular variation. *Genetics* **134**, 1289–1303 (1993).
- McDonald, J. H. & Kreitman, M. Adaptive protein evolution at the *Adh* locus in *Drosophila*. *Nature* **351**, 652–654 (1991).
- Jenkins, D. L., Ortori, C. A. & Brookfield, J. F. A test for adaptive change in DNA sequences controlling transcription. *Proc. R. Soc. Lond. B* **261**, 203–207 (1995).
- Egea, R., Casillas, S. & Barbadilla, A. Standard and generalized McDonald-Kreitman test: a website to detect selection by comparing different classes of DNA sites. *Nucleic Acids Res.* **36**, W157–W162 (2008).
- Sawyer, S. A. & Hartl, D. L. Population genetics of polymorphism and divergence. *Genetics* **132**, 1161–1176 (1992).
- Nielsen, R. Statistical tests of selective neutrality in the age of genomics. *Heredity* **86**, 641–647 (2001).
- Eyre-Walker, A. Changing effective population size and the McDonald-Kreitman test. *Genetics* **162**, 2017–2024 (2002).
- Charlesworth, J. & Eyre-Walker, A. The McDonald-Kreitman test and slightly deleterious mutations. *Mol. Biol. Evol.* **25**, 1007–1015 (2008).
- Eyre-Walker, A. & Keightley, P. D. Estimating the rate of adaptive molecular evolution in the presence of slightly deleterious mutations and population size change. *Mol. Biol. Evol.* **26**, 2097–2108 (2009).
- Fay, J. C., Wyckoff, G. J. & Wu, C. L. Testing the neutral theory of molecular evolution with genomic data from *Drosophila*. *Nature* **415**, 1024–1026 (2002).
- Ohta, T. Slightly deleterious mutant substitutions in evolution. *Nature* **246**, 96–98 (1973).
- David, J. R. & Capi, P. Genetic variation of *Drosophila melanogaster* natural populations. *Trends Genet.* **4**, 106–111 (1988).
- Begun, D. J. & Aquadro, C. F. African and North American populations of *Drosophila melanogaster* are very different at the DNA level. *Nature* **365**, 548–550 (1993).
- Tajima, F. Statistical methods to test for nucleotide mutation hypothesis by DNA polymorphism. *Genetics* **123**, 585–595 (1989).
- Smith, N. G. & Eyre-Walker, A. Adaptive protein evolution in *Drosophila*. *Nature* **415**, 1022–1024 (2002).
- Stoletzki, N. & Eyre-Walker, A. Estimation of the neutrality index. *Mol. Biol. Evol.* **28**, 63–70 (2011).
- Ayroles, J. F. et al. Systems genetics of complex traits in *Drosophila melanogaster*. *Nature Genet.* **41**, 299–307 (2009).
- Yamamoto, A. et al. Neurogenetic networks for startle-induced locomotion in *Drosophila melanogaster*. *Proc. Natl Acad. Sci. USA* **105**, 12393–12398 (2008).
- Harbison, S. T., Yamamoto, A. H., Fanara, J. J., Norga, K. K. & Mackay, T. F. C. Quantitative trait loci affecting starvation resistance in *Drosophila melanogaster*. *Genetics* **166**, 1807–1823 (2004).
- Pospisilik, J. A. et al. *Drosophila* genome-wide obesity screen reveals hedgehog as a determinant of brown versus white adipose cell fate. *Cell* **140**, 148–160 (2010).
- Bellen, H. J. et al. The BDGP gene disruption project: single transposon insertions associated with 40% of *Drosophila* genes. *Genetics* **167**, 761–781 (2004).
- The ModENCODE Consortium. Identification of functional elements and regulatory circuits by *Drosophila* modENCODE. *Science* **330**, 1787–1797 (2010).
- Charlesworth, B., Coyne, J. A. & Barton, N. H. The relative rates of evolution of sex chromosomes and autosomes. *Am. Nat.* **130**, 113–146 (1987).
- Sieberts, S. K. & Schadt, E. E. Moving toward a system genetics view of disease. *Mamm. Genome* **18**, 389–401 (2007).

Supplementary Information is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** This work was supported by National Institutes of Health grant GM 45146 to T.F.C.M., E.A.S. and R.R.H.A.; R01 GM 059469 to R.R.H.A., MCI BFU 2009-09504 to A.B., R01 GM 085183 to K.R.T., NHGRI U54 HG003273 to R.A.G.; and an award through the NVIDIA Foundation's "Compute the Cure" programme to D.M.

**Author Contributions** T.F.C.M., S.R. and R.A.G. conceived the project. T.F.C.M., S.R., A.B. and E.A.S. wrote the main manuscript. T.F.C.M., S.R., A.B., E.A.S., J.F.A., K.R.T., J.M.C., C.M.B. and D.M. wrote the Supplementary methods. M.M.M., C.B., K.P.B., M.A.C., L.C., L.D., Y.H., M.J., J.C.J., S.H.J., K.W.J., F. Lara, F. Lawrence, S.L.L., R.F.L., M.M., D.M.M., L.N., I.M., L.P., L.L.P., C.Q., J.G.R., S.M.R., L.T., K.C.W., Y.-Q.W., A.Y. and Y.Z. performed experiments. T.F.C.M., A.B., J.F.A., D.Z., S.C., M.M.M., J.M.C., M.F.R., M.B., D.C., R.S.L., A.M., C.M.B., K.R.T., D.M. and E.A.S. did the bioinformatics and data analysis. J.F.A., S.C., M.M.M., Z.H., P.L., M.R., J.R. and E.A.S. wrote the Methods and did the web site development. R.R.H.A. contributed resources.

**Author Information** Sequences have been deposited at the National Center for Biotechnology Information Short Read Archives (<http://www.ncbi.nlm.nih.gov/sra?term=DGRP>). Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). This paper is distributed under the terms of the Creative Commons Attribution-Non-Commercial-Share Alike licence, and is freely available to all readers at [www.nature.com/nature](http://www.nature.com/nature). The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at [www.nature.com/nature](http://www.nature.com/nature). Correspondence and requests for materials should be addressed to T.F.C.M. ([trudy\\_mackay@ncsu.edu](mailto:trudy_mackay@ncsu.edu)).



---

## 8. Índice de figuras

Figura 1.1. <i>Importancia relativa de los distintos tipos de selección natural propuestas en las principales teorías evolutivas desde de Darwin a la actualidad</i>	5
Figura 1.2. <i>Tipos de mutaciones</i>	7
Figura 1.3. <i>Esquema de la simulación de una población Wright-Fisher y su correspondiente árbol de coalescencia</i>	9
Figura 1.4. <i>Esquema que muestra la trayectoria de los alelos neutros y selectivos en una población</i>	15
Figura 1.5. <i>Modelo molecular de recombinación homóloga en eucariotas</i>	22
Figura 1.6. <i>Diseño experimental seguido para obtener las estimas de recombinación</i>	26
Figura 1.7. <i>Tasa de recombinación por brazo cromosómico y cruce</i>	27
Figura 1.8. <i>Tasa de conversión génica a lo largo de los brazos cromosómicos</i>	28
Figura 1.9. <i>La unidad de selección depende de la tasa de recombinación local</i>	29
Figura 1.10. <i>Interferencia de Hill-Robertson</i>	30
Figura 1.11. <i>Modelos de arrastre de la variación genética por selección</i>	31
Figura 1.12. <i>La paradoja de la variación genética</i>	35
Figura 1.13. <i>Esquema de la relación entre la diversidad nucleotídica y el censo efectivo de la población</i>	36
Figura 1.14. <i>Simulación del efecto de la selección ligada en la heterocigosidad de un locus</i>	37
Figura 1.15. <i>Correlación de las características del ciclo vital y la diversidad genética de distintas especies</i>	38
Figura 1.16. <i>Esquema de los cromosomas de D. melanogaster</i>	40
Figura 1.17. <i>Filogenia del género Drosophila</i>	41
Figura 1.18. <i>Dirección de la selección bajo dos contextos recombinacionales (<math>\geq 2cM/MB</math> y <math>&lt; 2cM/MB</math>)</i>	44
Figura 1.19. <i>Regímenes de selección en el genoma de D. melanogaster</i>	45
Figura 2.1. <i>Diseño experimental seguido para la obtención y secuenciación de las líneas DGRP</i>	52
Figura 2.2. <i>Pipeline para la estimación de la diversidad nucleotídica y la divergencia en la aproximación centrada en genes</i>	54
Figura 3.1. <i>Polimorfismo y divergencia en el genoma de D. melanogaster</i>	71
Figura 3.2. <i>Polimorfismo genómico de autosomas versus polimorfismo genómico corregido para el cromosoma X</i>	72
Figura 3.3. <i>Polimorfismo cromosómico de autosomas versus polimorfismo cromosómico corregido para el cromosoma X</i>	73
Figura 3.4. <i>Polimorfismo y divergencia de los sitios 4 veces degenerados</i>	76
Figura 3.5. <i>Polimorfismo en las distintas clases de sitios del genoma</i>	79
Figura 3.6. <i>Divergencia en las distintas clases de sitios del genoma</i>	79
Figura 3.7. <i>Razón <math>\pi_0/\pi_4</math> y <math>k_0/k_4</math> en distintas regiones del genoma</i>	81
Figura 3.8. <i>Patrón de recombinación a lo largo de cada brazo cromosómico</i>	86
Figura 3.9. <i>Regresión local suavizada entre el polimorfismo y la recombinación</i>	90
Figura 3.10. <i>Mapa de bloques de selección ligada (LSB) del genoma de D. melanogaster cartografiado a partir de ventanas de 100kb</i>	94
Figura 3.11. <i>Tasa de recombinación a lo largo del cromosoma superpuesta el mapa de LSB</i>	96
Figura 3.12. <i>Distribución de la razón <math>c/\mu</math> observada</i>	101

---

Figura 3.13. Simulación de los valores de $\alpha$ y $\theta$ en función de la razón $c/\mu$ _____	102
Figura 3.14. Desequilibrio de ligamiento, $r^2$ , en los bloques de selección ligada (LSB) _____	104
Figura 3.15. Mapa de LSB en diferentes tamaños de ventana_____	115
Figura 4.1. Interpretación de los niveles de diversidad nucleotídica bajo (a) la teoría neutralista y (b) la selección ligada recurrente (draft genético) _____	125
Figura 4.2. Esquema ilustrativo del efecto del tamaño de ventana sobre la estimación del valor umbral de recombinación _____	128
Figura 4.3. Gráfico que relaciona la distancia entre dos marcadores y su probabilidad que recombinen por conversión génica o por entrecruzamiento. _____	130

---

## 9. Índice de tablas

Tabla 1.1. <i>Estimas de la tasa de mutación por sitio nucleotídico</i>	7
Tabla 1.2. <i>Métodos empleados para la detección de las huellas que deja la selección natural en las secuencias de DNA</i>	12
Tabla 1.3. <i>Estimadores de los regímenes de selección derivados de la extensión del MKT</i>	16
Tabla 1.4. <i>Estimas de la proporción de la evolución adaptativa (<math>\alpha</math>) en diferentes especies y regiones genómicas</i>	18
Tabla 1.5. <i>Estima del porcentaje de alelos bajo selección negativa en distintas especies y regiones funcionales</i>	20
Tabla 1.6. <i>Estimas de la tasa de recombinación en distintas especies</i>	23
Tabla 1.7. <i>Estimadores del desequilibrio de ligamiento (LD)</i>	24
Tabla 1.8. <i>Relación entre la recombinación y las medidas de diversidad y divergencia</i>	33
Tabla 1.9. <i>Proyectos de resecuenciación en distintos organismos modelo</i>	43
Tabla 3.1. <i>Polimorfismo y divergencia en el genoma de D. melanogaster</i>	68
Tabla 3.2. <i>Comparativa de las distintas estimas de la diversidad nucleotídica en poblaciones norteamericanas</i>	69
Tabla 3.3. <i>Población norteamericana versus población africana</i>	71
Tabla 3.4. <i>Polimorfismo y divergencia en autosomas y el cromosoma X</i>	77
Tabla 3.5. <i>ANOVA del polimorfismo sinónimo y variables genómicas como la recombinación y mutación</i>	83
Tabla 3.6. <i>Correlación entre las distintas estimas de recombinación a lo largo del genoma</i>	85
Tabla 3.7. <i>Descripción estadística de las distintas estimas de recombinación por brazo cromosómico</i>	87
Tabla 3.8. <i>Correlación de la recombinación con el polimorfismo y con la divergencia en cada brazo cromosómico y el total del genoma</i>	88
Tabla 3.9. <i>Regresión lineal versus regresión segmentada entre el polimorfismo y la recombinación</i>	89
Tabla 3.10. <i>Valores estimados del umbral de recombinación</i>	91
Tabla 3.11. <i>Proporción de sitios de los LSB y los NLSB</i>	93
Tabla 3.12. <i>Número y tamaño de los LSB</i>	95
Tabla 3.13. <i>Número y proporción de genes en los LSB estimados</i>	97
Tabla 3.14. <i>Diferencias entre el número de genes de los LSB y NLSB</i>	98
Tabla 3.15. <i>Descripción estadística de la razón <math>c/\mu</math></i>	99
Tabla 3.16. <i>Simulación de los valores de <math>\alpha</math> y <math>\theta</math> en función de la razón <math>c/\mu</math></i>	103
Tabla 3.17. <i>Diferencias en <math>r^2</math> entre LSB y NLSB</i>	105
Tabla 3.18. <i>Correlación entre el polimorfismo y la recombinación en los LSB y NLSB</i>	106
Tabla 3.19. <i>Correlación del polimorfismo y la divergencia en los LSB y NLSB</i>	107
Tabla 3.20. <i>Estimas de selección en el genoma de D. melanogaster</i>	109
Tabla 3.21. <i>Estimas de selección en los LSB de D. melanogaster</i>	110
Tabla 3.22. <i>Diferencias de las estimas de selección en los LSB y NLSB</i>	110
Tabla 3.23. <i>Estimas de selección en los LSB tras excluir las regiones centroméricas</i>	111
Tabla 3.24. <i>Diferencias de las estimas de selección en los LSB y NLSB tras excluir las regiones centroméricas</i>	112
Tabla 3.25. <i>Relación entre el polimorfismo y la recombinación para distintos tamaños de ventana</i>	112

---

Tabla 3.26. <i>Correlaciones polimorfismo-recombinación y polimorfismo-divergencia en los LSB en ventanas de 10kb</i>	116
Tabla 3.27. <i>Estimas de selección en los LSB en ventanas de 10kb</i>	117
Tabla 3.28. <i>Estimas de selección en los LSB en ventanas de 10kb tras excluir las regiones centroméricas</i>	117

---

## 10. Índice de cuadros

Cuadro 1. Mutación _____	6
Cuadro 2. La teoría de la coalescencia _____	9
Cuadro 3. Cartografía genómica de la recombinación a alta resolución en <i>D. melanogaster</i> , c _____	26
Cuadro 4. <i>Draft</i> genético _____	37
Cuadro 5. Genómica de poblaciones de las secuencias DGRP _____	43



---

## 11. Índice de tablas suplementarias

Tabla suplementaria 1. <i>Estimas de diversidad y divergencia en el genoma de D. melanogaster</i>	166
Tabla suplementaria 2. <i>Mapa de bloques de selección ligada</i>	181
Tabla suplementaria 3. <i>Regiones centroméricas y teloméricas</i>	220
Tabla suplementaria 4. <i>Pruebas de significación para la existencia de diferencias entre brazos cromosómicos de las estimas de selección</i>	221



---

## AGRADECIMIENTOS

Gracias al Dr. Antonio Barbadilla por brindarme la oportunidad de llevar a cabo esta investigación, por enseñarme la seguridad y el respaldo que proporciona tener conocimientos teóricos a la hora de defender tus ideas científicas. Por sus maravillosas correcciones que hacen que los textos que escribo adquieran otra categoría.

A David, Miquel, Yolanda y Nuria por las conversaciones *freak* genética, de cómo arreglar el mundo o simplemente por reírnos juntos, algo tan necesario durante los años que ha durado la tesis. Lo echaré todo ello de menos. Querría agradecer especialmente a David las horas dedicadas a meditar, pensar y solucionar los problemas teóricos y técnicos surgidos a lo largo de esta tesis.

A la Dra. Josefa Gonzalez, que aunque no haya participado directamente en el trabajo, quería agradecerle su apoyo profesional y personal en la parte final de esta tesis.

A Carlos: Gracias por ser un resuelve problemas. Gracias por impulsarme a conseguir todo lo que quiero. Gracias por animarme en los días oscuros relativizando todos los problemas a lo más absurdo. Gracias por ser práctico y enseñarme y ayudarme con la programación, informática y diseño gráfico que he necesitado a lo largo de estos años.

Amari eta Aitari, hasi aurretik nigan konfidantza itxua izateagatik. Amona Maite eta Amona Antxoniri, zuen bizipoza eta maitasunagatik. Eskerrik asko, nere arrakasta guztiak zuenak ere bihurtzeagatik. Horregatik tesi hau zuena ere bada.

Eskerrik asko Mikele eta Patxiri zuen laguntzagatik eta babesagatik.

Azkenik eta bereziki Marc-i. Tesi honi eman dizkiodan azken orduak zuri kendu dizkizudalako.