

*Evaluación del desequilibrio gamético
dependiente de marcador en el brazo corto
del cromosoma 11 humano.
Estudio de asociación con la otosclerosis*

*Tesis Doctoral
Laura Rodríguez Murillo*



FACULTAD DE BIOLOGÍA
Departamento de Genética

EVALUACIÓN DEL DESEQUILIBRIO GAMÉTICO
DEPENDIENTE DE MARCADOR EN EL BRAZO CORTO
DEL CROMOSOMA 11 HUMANO.
ESTUDIO DE ASOCIACIÓN CON LA OTOSCLEROSIS

MEMORIA PRESENTADA POR:
LAURA RODRÍGUEZ MURILLO
PARA OPTAR AL GRADO DE DOCTORA EN BIOLOGÍA
Santiago de Compostela, 2005

EL DR. CARLOS ZAPATA BABÍO, CATEDRÁTICO DEL DEPARTAMENTO DE GENÉTICA DE LA UNIVERSIDAD DE SANTIAGO DE COMPOSTELA, Y EL DR. CARLOS MARTIN MARTIN, PROFESOR TITULAR DEL DEPARTAMENTO DE DERMATOLOGÍA Y OTORRINOLARINGOLOGÍA DE LA UNIVERSIDAD DE SANTIAGO DE COMPOSTELA

INFORMAN:

Que el trabajo titulado “EVALUACIÓN DEL DESEQUILIBRIO GAMÉTICO DEPENDIENTE DE MARCADOR EN EL BRAZO CORTO DEL CROMOSOMA 11 HUMANO. ESTUDIO DE ASOCIACIÓN CON LA OTOSCLEROSIS”, que presenta LAURA RODRÍGUEZ MURILLO para la obtención del Grado de Doctora en Biología, ha sido realizado bajo su dirección, que lo consideran concluido y autorizan su presentación al tribunal calificador.

Y para que así conste, se firma el presente informe en Santiago de Compostela, a 29 de Abril de dos mil cinco.

Fdo: Carlos Zapata Babío

Fdo.: Carlos Martín Martín

Este trabajo ha sido financiado mediante los Proyectos de Investigación: BMC2001-3444 concedido por el Ministerio de Ciencia y Tecnología al Dr. Carlos Zapata Babío, PGIDIT03PXIC20003PN concedido por la *Xunta de Galicia* al Dr. Carlos Zapata Babío, PGIDIT04RMA261014PR concedido por la *Xunta de Galicia* al Dr. Carlos Zapata Babío y FIS 98/160 concedido por el Ministerio de Sanidad y Consumo al Dr. Carlos Martín Martín. Asimismo, su realización se ha visto facilitada gracias al disfrute de una *Bolsa de Terceiro Ciclo* y una *Bolsa Predoutoral* correspondiente al *Plan Galego de Investigación e Desenvolvemento* concedidas por la *Dirección Xeral de Investigación e Desenvolvemento* de la *Xunta de Galicia*; y a una *Bolsa de Cooperación en Contratos e Convenios* concedida por la Universidad de Santiago de Compostela.

El contenido de esta Memoria no podrá reproducirse, ya sea todo o en parte, bajo ningún tipo de soporte, sin la autorización expresa del Dr. J. Carlos Zapata Babío.

Y para que conste a los efectos oportunos, firmamos el presente informe en Santiago de Compostela, a 29 de Abril de dos mil cinco.

Fdo.: J. Carlos Zapata Babío

Fdo.: Laura Rodríguez Murillo

A la madre naturaleza

Agradecimientos

Los que me conocéis bien sabéis que soy más bien de pocas palabras, pero lo poco que digo es muy sincero y espero que estas líneas que vienen a continuación sirvan para dar las gracias a aquellos que de una u otra manera habéis contribuido al desarrollo de este trabajo dentro y fuera del laboratorio.

Ante todo, debo agradecer sinceramente todos estos años de investigación a mis directores de Tesis Doctoral. Al Dr. Carlos Zapata, por haberme dado la oportunidad de realizar esta investigación, por su ayuda prestada en todo momento y por sus más que valiosos consejos durante la redacción de este trabajo. Agradecer igualmente al Dr. Carlos Martín la vertiente médica en la dirección de este trabajo y el que siempre haya tenido un momento para recibirme sin previo aviso.

Mil gracias a todos los que de alguna manera forman o formaron parte de este laboratorio y que compartieron conmigo su tiempo y conocimientos. Conchi, Esther, gracias por vuestro apoyo, por haberme acogido como lo hicisteis y por los viajes compartidos. Amparo, te mereces todo lo mejor en el mundo. No sabes cómo he disfrutado de tus conversaciones y tu cariño a lo largo de estos años, eres un sol. Roberto, gracias por esos caramelos de vitamina C. Ana “Boente” por nuestras charlas viajeras de los Miércoles.

Sobre todo a mis amigas Eunice, Tati, Cristina y Olga, porque sabéis como nadie lo difícil que ha sido todo para mí, porque sois un apoyo fundamental y siempre habéis confiado en mí, muchas gracias chicas.

A los compañeros de piscina, Alex, María, por el cariño que me disteis en un momento en que más lo necesitaba y porqué no, por esas salidas nocturnas. Julián, gracias por esos desayunos con bocata compartido. A todos, por los madrugones (yo siempre tarde) y por los partidillos de waterpolo, chicos la piscina sin vosotros es muy aburrida.

Agradecimientos

Gracias a todas las partes de mi familia porque, al fin y al cabo, sin vosotros yo no estaría aquí.

Y muchas gracias Christian porque tú has vivido la última etapa de este trabajo tanto como yo, ahora toca disfrutarlo juntos.

Abreviaturas

cM	centiMorgan
DF	Distancia Física
DG	Desequilibrio Gamético
DGGE	Electroforesis en geles con gradiente de desnaturalización (<i>Denaturing Gradient Gel Electrophoresis</i>)
DNA	Ácido desoxirribonucleico
EDTA	Ácido Etilendiaminotetraacético Sal Disódica (<i>Ethylenediaminetetraacetic Acid Disodium Salt</i>)
E-M	Expectación-Maximización
FRE	Frecuencia de Recombinación Efectiva
g.l.	Grados de libertad
HapMap	Mapa de Haplotipos (<i>Haplotype Map</i>)
h_e	Heterozigosis esperada
h_o	Heterozigosis observada
H-W	Hardy-Weinberg
kb	Kilobases
Mb	Megabases
NCBI	<i>National Center for Biotechnology Information</i>
NTSYS	<i>Numerical Taxonomy and Multivariate Analysis System</i>
P	Probabilidad
pb	Pares de bases
PCR	Reacción en cadena de la polimerasa (<i>Polimerase Chain Reaction</i>)
RFLPs	Polimorfismos de longitud de fragmentos de restricción (<i>Restriction Fragment Length Polymorphisms</i>)
rpm	Revoluciones por minuto
SE	Error estándar (<i>Standard Error</i>)
SNPs	Polimorfismos de Nucleótidos Sencillos (<i>Single Nucleotide Polymorphisms</i>)
Tris	<i>Trishydroxymethyl-aminomethane</i>
TSC	<i>The SNP Consortium</i>

Índice

1. RESUMEN/SUMMARY.....	1
2. INTRODUCCIÓN.....	5
1. El DG en la detección de genes implicados en enfermedades.....	12
2. El DG dependiente de marcador.....	16
3. El caso de la otosclerosis.....	30
4. Objetivos.....	36
3. MATERIAL Y MÉTODOS.....	39
I. EVALUACIÓN DEL DG DEPENDIENTE DE MARCADOR EN 11P.....	41
1. Marcadores genéticos analizados.....	41
1.1. Distancias físicas (DF) y frecuencias de recombinación efectivas (FRE) entre los loci.....	45
1.1.1. DF entre los marcadores genéticos analizados.....	45
1.1.2. FRE entre pares de SNPs y entre pares de microsatélites.....	49
2. Genotipado.....	54
2.1. Genotipado de los loci microsatélites.....	54
2.2. Genotipado de los SNPs.....	54
2.2.1. Material biológico.....	54
2.2.2. Extracción de DNA.....	55
2.2.3. Genotipado de los SNPs analizados mediante PCR-RFLP....	55
2.2.3.1. Amplificación por la PCR de los SNPs.....	55
2.2.3.2. Amplificación del locus HRAS-Nsil.....	59
2.2.3.3. Digestión de los productos de amplificación para los SNPs.....	60
2.2.3.4. Resolución de los genotipos: Electroforesis.....	62
2.2.3.4.1. Preparación del gel.....	62
2.2.3.4.2. Sistema electroforético.....	63
2.2.3.4.3. Tinción del gel.....	64
2.2.4. Genotipado de los SNPs analizados mediante DGGE.....	65

2.2.4.1. Amplificación por PCR de los SNPs analizados mediante DGGE.....	66
2.2.4.2. DGGE.....	68
2.2.4.2.1. Preparación del gel de electroforesis.....	69
2.2.4.2.2. Carga del gel y migración electroforética.....	70
2.2.4.2.3. Visualización de los resultados.....	70
3. Análisis de los datos.....	71
3.1. Análisis de la variabilidad y estructura genética poblacional.....	71
3.1.1. Variabilidad genética.....	71
3.1.2. Magnitud y significación de las desviaciones de las proporciones Hardy-Weinberg.....	72
3.1.2.1. Magnitud de las desviaciones.....	72
3.1.2.2. Significación de las desviaciones.....	74
3.2. Estimación del desequilibrio gamético (DG).....	75
3.2.1. Estimación de las frecuencias haplotípicas.....	75
3.2.2. Estimación de la intensidad del DG.....	77
3.2.2.1. Estimación de la intensidad del DG entre loci bialélicos (SNPs).....	77
3.2.2.2. Estimación de la intensidad del DG entre pares de loci multialélicos (microsatélites).....	79
3.2.2.2.1. Estimación de la intensidad de los DG interalélicos.....	79
3.2.2.2.2. Estimación de la intensidad de los DG globales de signo positivo.....	81
3.2.3. Significación del DG.....	82
3.2.3.1. Significación del DG entre pares de loci bialélicos (SNPs).....	82
3.2.3.2. Significación del DG entre pares de loci multialélicos (microsatélites).....	83
3.2.3.2.1. Significación del DG interalélico positivo.....	83
3.2.3.2.2. Significación del DG global positivo.....	83
3.3. Relación del DG con la FRE y la DF.....	84

II. ESTUDIO DE ASOCIACIÓN DE LA OTOSCLEROSIS CON LOS GENES DEL COLÁGENO TIPO I, COL1A1 Y COL1A2.....	86
1. Material biológico.....	86
1.1. Extracción de DNA.....	86
2. Marcadores genéticos analizados.....	87
3. Amplificación por la PCR de los marcadores genéticos.....	89
4. Digestión de los productos de amplificación para los SNPs.....	92
5. Resolución de los genotipos: Electroforesis.....	92
6. Análisis de los datos.....	94
6.1. Análisis de la variabilidad y estructura genética poblacional.....	94
6.2. Estimación de las frecuencias haplotípicas.....	95
6.3. Análisis de asociación.....	96
6.3.1. Estimación de la intensidad de la asociación y su significación.....	97
4. RESULTADOS.....	99
I. EVALUACIÓN DEL DG DEPENDIENTE DE MARCADOR EN 11P.....	101
1. Estudio del DG entre pares de SNPs.....	101
1.1. Interpretación de los patrones electroforéticos de los SNPs.....	101
1.1.1. SNPs analizados mediante PCR-RFLP.....	101
1.1.2. SNPs analizados mediante DGGE.....	105
1.2. Análisis de la variabilidad y la estructura poblacional.....	106
1.3. Análisis del DG.....	110
1.3.1. DG entre pares de SNPs.....	111
1.3.2. Relación del DG con la FRE y la DF.....	115
2. Estudio del DG entre pares de loci microsatélites.....	121
2.1. Análisis de la variabilidad y estructura genética poblacional.....	121
2.2. Análisis del DG.....	129
2.2.1. DG interalélicos.....	130
2.2.2. Relación del DG interalélico con el tamaño y frecuencia de los alelos.....	136
2.2.3. DG globales positivos.....	137

2.2.4. Relación del DG con la FRE y la DF.....	144
2.2.4.1. Relación del DG interalélico con la FRE y la DF.....	144
2.2.4.2. Relación del DG global con la FRE y la DF.....	154
II. ESTUDIO DE ASOCIACIÓN DE LA OTOSCLEROSIS CON LOS GENES DEL COLÁGENO TIPO I, COL1A1 Y COL1A2.....	163
1. Interpretación de los patrones electroforéticos.....	163
2. Análisis de la variabilidad y estructura genética poblacional.....	167
3. Análisis de asociación.....	174
5. DISCUSIÓN.....	179
I. EVALUACIÓN DEL DG DEPENDIENTE DE MARCADOR EN 11P.....	181
II. ESTUDIO DE ASOCIACIÓN DE LA OTOSCLEROSIS CON LOS GENES DEL COLÁGENO TIPO I, COL1A1 Y COL1A2.....	207
6. CONCLUSIONES.....	213
7. BIBLIOGRAFÍA.....	219

1. Resumen/Summary

La información del desequilibrio gamético (DG) de *background* es relevante para la realización de un correcto diseño e interpretación de los estudios de localización de genes de enfermedades mediante DG. Estos estudios se han realizado mediante marcadores genéticos con dinámicas evolutivas diferentes como son los SNPs y los microsatélites. Por lo tanto, la cuantificación del DG puede verse perturbada por la existencia de DG dependiente de marcador, de manera que se requiere averiguar qué fracción del DG de *background* es debida a las características propias de cada marcador. En el presente trabajo se ha realizado una evaluación del DG dependiente de marcador con 18 SNPs y 22 microsatélites localizados a lo largo de una extensa región anónima del genoma humano, como es el brazo corto del cromosoma 11. El DG se comparó entre pares de SNPs y entre pares de microsatélites en dos muestras independientes de 405 individuos de población gallega. Los resultados mostraron una mayor cantidad de DG entre pares de microsatélites que entre pares de SNPs. Sin embargo, en general, la magnitud de las asociaciones fue superior entre pares de SNPs, que entre pares de microsatélites. El DG entre pares de microsatélites reveló un patrón de DG dependiente de la frecuencia y tamaño de los alelos implicados. Este patrón de DG dependiente de alelo confirma resultados previos obtenidos en las regiones cromosómicas 11p15 y Xq25-28, y presenta una explicación basada en la alta tasa y compleja dinámica mutacional de los microsatélites. El DG se extiende hasta distancias muy inferiores entre pares de SNPs que entre pares de microsatélites. Así, la mayoría de los pares de SNPs en DG están separados por menos de 0,5 cM. Sin embargo, la distancia entre pares de microsatélites en DG es próxima a la segregación independiente (42 cM). Además, la relación de la intensidad del DG con la frecuencia de recombinación entre pares de SNPs sigue una correlación negativa significativa. Por el contrario, esta tendencia no se observa entre pares de microsatélites, a no ser que consideremos exclusivamente los alelos más frecuentes. La diferente dinámica mutacional de los SNPs y los microsatélites parece ser el principal factor determinante para las diferencias observadas en los patrones de DG con uno y otro tipo de marcador. Los resultados sugieren que los SNPs son mejores marcadores que los microsatélites del DG de *background* en el genoma. Así, el DG dependiente de marcador representa una fracción importante del DG de *background* a lo largo del genoma. Por esta razón, será relevante su cuantificación para otras regiones del genoma con el fin de realizar apropiadamente el diseño e interpretación de los estudios de localización de genes de enfermedades mediante DG.

La otosclerosis es una de las causas más comunes de pérdida de audición entre adultos de origen caucasoide. Los genes COL1A1 y COL1A2 se han propuesto como genes candidatos implicados en la otosclerosis. En el presente trabajo se presenta un estudio independiente de asociación entre la otosclerosis y polimorfismos en los genes COL1A1 y COL1A2 con 100 individuos afectados y 100 controles sanos, todos provenientes de la población gallega. Específicamente, se estudiaron dos marcadores en COL1A1 que habían sido asociados con la otosclerosis en un estudio previo, y seis marcadores en COL1A2. Los análisis de asociación se basaron en la comparación de las frecuencias alélicas, genotípicas y haplotípicas de dos loci entre individuos afectados y controles. Los resultados no aportan ninguna evidencia para una posible implicación de los genes COL1A1 y COL1A2 en la etiología de la otosclerosis.

Summary

Information on the amount and distribution of gametic disequilibrium (GD) across the human genome (background disequilibrium) is relevant for successful localization of disease genes by GD. These studies have been done with genetic markers with very different evolutionary dynamics as SNPs and microsatellites. It follows that quantification of the amount of background GD can be seriously confounded by the occurrence of marker-dependent disequilibrium. Here, we report a study aimed at assessing the marker-dependent GD with 18 SNPs and 22 microsatellites across an extensive anonymous region of the human genome, as it is the short arm of chromosome 11. GD between SNP pairs and between microsatellite pairs was compared in two independent samples of 405 individuals, from Galician population. The results show that the amount of GD between microsatellite pairs is greater than for SNPs. However, in general, the strength of associations is higher for SNP pairs than for microsatellite pairs. Microsatellite pairs revealed pairwise GD patterns dependent on allele frequency and size. This allele-dependent GD pattern confirms previously published results in 11p15 and Xq25-28, and it can be explained on the basis of the high mutation rate and complex mutation dynamics of microsatellites. Moreover, GD between SNP pairs extends over shorter regions than between microsatellite pairs. Thus, most SNP pairs in GD are separated by less than 0,5 cM and the distance between microsatellite pairs almost reaches the limit of independent assortment (42 cM). In addition, the relationship of the strength of GD with recombination frequency between SNP pairs follows a significant negative correlation. However, this relation is not found between microsatellite pairs, unless we consider only the most frequent alleles. The different mutational dynamics of SNPs and microsatellites seems to be the main reason for the observed differences in GD patterns measured with each type of marker. The results suggest that the SNPs are better markers than microsatellites for description of background GD along the genome. In conclusion, marker-dependent GD represents an important portion of the background GD along the genome; consequently, its quantification for other regions of the genome will be relevant for designing and interpreting experiments of localization of disease genes by GD mapping.

Otosclerosis is one of the most common causes of hearing loss in white adults. The COL1A1 and COL1A2 genes have been proposed as candidate genes in the development of otosclerosis. Here, we present an independent study of association between COL1A1 and COL1A2 gene polymorphisms and otosclerosis, in a case-control sample with 100 affected and 100 control individuals from Galician population. Specifically, we study two markers in COL1A1 previously reported to be associated with otosclerosis, and six markers in COL1A2. We performed diverse association analyses based on alleles, genotypes, and two-locus haplotypes. We found no evidence supporting the putative link of COL1A1 and COL1A2 genes with otosclerosis.

2. Introducción

El término “genoma”, acuñado en el año 1920 por el botánico alemán, Hans Winkler, para designar el conjunto de cromosomas haploides de eucariotas; se podría definir actualmente como algo más que la suma de sus partes, esencialmente debido a las interacciones estructurales, funcionales y evolutivas que ocurren entre sus diferentes regiones (BERNARDI 1995). Cada genoma es único y su secuencia de nucleótidos difiere entre individuos aproximadamente en 1 nucleótido cada 300 pb (LI Y SADLER 1991; PRZEWORSKI *et al.* 2000; KRUGLYAK Y NICKERSON 2001; FORTON *et al.* 2005). Este polimorfismo del DNA influye en características individuales como la apariencia física, susceptibilidad a enfermedades (RISCH Y MERIKANGAS 1996; KRUGLYAK 1997; JORDE 2000) o respuesta a tratamientos farmacológicos (ROSES 2001; WILSON *et al.* 2001; GOLDSTEIN *et al.* 2003). El polimorfismo del DNA también representa un registro fósil de la historia y la estructura de las poblaciones ancestrales (REICH *et al.* 2002).

El estudio de las asociaciones no al azar entre alelos de loci distintos o desequilibrio gamético (DG) es fundamental a la hora de estudiar esa variación interindividual e interpoblacional y permite abordar la tarea de la localización de genes implicados en enfermedades (LANDER y SCHORK 1994; RISCH y MERIKANGAS 1996; KRUGLYAK 1999; JORDE 2000; PRITCHARD y PRZEWORSKI 2001; ARDLIE *et al.* 2002; BOTSTEIN y RISCH 2003; WALL y PRITCHARD 2003). Los patrones de DG pueden ser también de utilidad para inferir la distribución de los puntos calientes de recombinación en el cromosoma (DAWSON *et al.* 2002; GABRIEL *et al.* 2002; PHILLIPS *et al.* 2003). Además, puede ser de gran ayuda para el estudio de fenómenos de la historia evolutiva de las poblaciones como crecimiento poblacional, mutaciones, deriva genética, cuellos de botella, movimientos migratorios y mezcla (TISHKOFF *et al.* 1996; HEDRICK 2005; REICH *et al.* 2001; NORDBORG y TAVARE 2002; CAMP *et al.* 2005; ZHANG *et al.* 2004), así como para el estudio de la selección natural (TISHKOFF *et al.* 2001; SABETI *et al.* 2002). Por otro lado, el DG puede ayudar a desentrañar la estructura genética de las poblaciones actuales (PRITCHARD y ROSENBERG 1999).

Tradicionalmente, el estudio del DG jugó un papel fundamental a la hora de decidir entre la teoría genética clásica, donde el gen es la unidad de selección; o la teoría genética multiloci que otorga relevancia a las interacciones que puedan existir entre los loci (FRANKLIN Y LEWONTIN 1970; LEWONTIN 1974). Históricamente, se suponía que la selección natural actuaba en genes individuales, de manera que la evolución se podía describir en términos de cambios en las frecuencias alélicas (CLEGG 1984). Con la maduración en este campo, se desarrolló la idea de que la dinámica evolutiva de cualquier gen particular podría estar influenciada por asociaciones estadísticas con genes en loci distintos. De esta manera surgió el estudio de la genética multiloci como una teoría más completa que evaluaba las consecuencias de la transmisión de estas asociaciones (ASMUSEN Y CLEGG 1981). Así, si se demostrase que los valores de DG no eran relevantes en las poblaciones cobraría fuerza la idea del gen como unidad de selección. Si, por el contrario, el DG resultara ser relevante en las poblaciones naturales, la teoría de un solo locus dejaría de ser válida. De este modo, para un conocimiento más profundo de la estructura genética de las poblaciones, se necesitaría conocer, además de las frecuencias alélicas de los loci individuales, el grado en el que esos loci se asociaban no al azar y la extensión de esas asociaciones.

Los primeros estudios teóricos encaminados a estudiar el DG en las poblaciones se centraron en el sistema más sencillo, dos loci con dos alelos. En estas investigaciones examinaron el DG bajo un modelo simétrico de viabilidad de dos loci con dos alelos (LEWONTIN Y KOJIMA 1960; BODMER Y PARSONS 1962). Estos autores concluyeron que el DG sólo estaba presente en las poblaciones cuando se cumplían condiciones especiales como fuerte epistasis respecto al valor de recombinación o estrecho ligamiento. Pero en trabajos posteriores se observó que las condiciones para la existencia de un DG estable entre pares de loci, podrían ser menos astringentes en el contexto de un sistema con mayor número de loci (LEWONTIN 1964; FRANKLIN y LEWONTIN 1970; SLATKIN 1972). En estos trabajos de simulación por ordenador, las condiciones para obtener un DG estable se relajaban con respecto a la teoría de dos loci. Estos resultados sugerían que dos loci adyacentes se mantendrían en mayor DG que lo predicho por la teoría de dos

loci, cuando estuvieran incluidos en una cadena con otros loci interactuando con ellos (efecto de empujamiento). Además, loci lejanos entre sí en el cromosoma podrían mantener un DG estable debido a asociaciones con loci situados entre ellos (efecto acumulativo) (FRANKLIN y LEWONTIN 1970).

Los estudios de DG para sistemas genéticos multiloci parecían indicar que éste debería ser un fenómeno general en las poblaciones. Pero los primeros análisis experimentales en los años 60 con poblaciones naturales de *Drosophila* sugerían que el DG era escaso en la naturaleza incluso entre loci proteicos fuertemente ligados (HEDRICK *et al.* 1978; BARKER 1979; LEWONTIN 1985; HEDRICK 2005). Estos resultados fueron sorprendentes dada la gran cantidad de factores poblacionales, además de la selección, que potencialmente podrían estar generando DG, deriva génica, efectos fundadores, cuellos de botella, migración, mutación, *hitchhiking* (HEDRICK *et al.* 1978). Para genes neutrales en poblaciones de tamaño finito la deriva genética puede ser la responsable de crear asociaciones entre alelos de loci distintos (HILL y ROBERTSON 1968; OHTA y KIMURA 1969; SLATKIN 1994). Así, incluso con un tamaño poblacional grande en la actualidad, el DG puede estar presente debido a su procedencia de una pequeña población fundadora o por la existencia en el pasado de un cuello de botella en la población (REICH *et al.* 2001; PTAK *et al.* 2004). Así mismo, la migración entre poblaciones puede generar asociaciones no aleatorias entre loci distintos cuando sus frecuencias alélicas son diferentes (HEDRICK *et al.* 1978). El efecto de la subdivisión o estratificación poblacional es esencialmente el mismo que la migración, de manera que si los grupos poblacionales mantienen poco flujo génico habrá DG debido a sus diferentes frecuencias alélicas (NEI y LI 1973; LI y NEI 1974; OHTA 1982; WILSON y GOLDSTEIN 2000).

Otro factor que puede generar DG es la mutación (OHTA y KIMURA 1969, 1970; HEDRICK *et al.* 1978). Probablemente la mutación por sí misma genera cantidades insignificantes de DG, sin embargo, mutación, migración y recombinación son las fuentes de nuevos tipos gaméticos en la población. Nuevos gametos, producidos por mutación, pueden aumentar su frecuencia por efecto de

la deriva genética o por selección lo que puede generar un DG subsiguiente (HEDRICK *et al.* 1978; HEDRICK 2005). Por último, cabe citar un fenómeno que puede generar DG entre loci neutrales para la selección. Este fenómeno es el arrastre genético o *hitchhiking*. Mediante este fenómeno, la selección actuando sobre un locus, puede cambiar las frecuencias alélicas de otro locus neutral que esté ligado al locus que está siendo seleccionado (KOJIMA y SCHAFER 1967; NACHMAN 2001; PRZEWORSKI 2002).

Por otro lado, todos los factores que disminuyan el número de dobles heterocigotos sobre los que puede actuar la recombinación rompiendo las asociaciones, retardarán el decaimiento del DG. Entre estos factores están el ligamiento, las inversiones cromosómicas y las desviaciones del apareamiento al azar. Además, la consanguinidad actúa de manera similar al ligamiento retardando el decaimiento del DG (ALLARD *et al.* 1968; KARLIN 1969; WEIR *et al.* 1972; WEIR y COCKERHAM 1973; HEDRICK *et al.* 1978). Por último, cabe señalar los factores que aceleran el decaimiento del DG que son, principalmente, la conversión génica, la mutación recurrente y la recombinación actuando sobre los dobles heterocigotos. Los dos primeros actúan a cortas distancias y en baja frecuencia, pero a largas distancias la recombinación es preponderante.

Tras la consideración de todos los fenómenos que podrían generar DG, se requería la búsqueda de causas a la escasez de DG observado en las poblaciones naturales. En un principio se trabajó con la posibilidad de que las técnicas electroforéticas utilizadas con loci alozímicos no estuvieran detectando toda la variabilidad existente. Además existía la idea de que la falta de asociaciones fuese debida a las enormes distancias de mapa existentes entre los loci estudiados. Pero con la llegada de los polimorfismos de DNA se dispuso de una batería de marcadores moleculares lo suficientemente cercanos en el cromosoma como para mostrar DG, además de ser la fuente misma de la variabilidad. A pesar de esto, los primeros estudios en DNA con RFLPs (*Restriction Fragment Length Polymorphisms*) dieron resultados semejantes a los anteriores sugiriendo que el DG era un fenómeno escaso en las poblaciones naturales (HEDRICK *et al.* 1978;

BARKER 1979; LEWONTIN 1985). Incluso más recientemente el DG parecía ser escaso en el DNA a distancias mayores de 0,1 cM (~ 0,1 Mb) (BODMER 1986; KRUGLYAK 1999; OTT 2000).

La falta de relación funcional entre los loci analizados se postuló también como una causa de la ausencia de DG en los estudios experimentales. Recientemente, ZAPATA *et al.* (2002) en un estudio con 15 loci proteicos distribuidos a lo largo del tercer cromosoma de *Drosophila melanogaster*, encontraron DG en distancias de hasta 15 – 20 cM entre loci relacionados funcionalmente. Sus resultados apoyaban la hipótesis de interacciones epistáticas en fitness actuando para determinar la cantidad y distribución del DG a lo largo del tercer cromosoma de *D. melanogaster*.

Otra de las posibles causas de la falta de detección de DG en las poblaciones naturales parecía apuntar a la escasa potencia estadística de los tests utilizados para detectar las asociaciones (BROWN 1975; HEDRICK *et al.* 1978; HASTINGS 1984, 1985; ZAPATA y ALVAREZ 1992). ZAPATA y ALVAREZ (1987, 1992) realizaron un metaanálisis de un conjunto de estudios de DG entre pares de loci proteicos con especies de *Drosophila* y encontraron que el valor promedio de potencia en la mayoría de los trabajos era tan sólo de un 20 %. La potencia estadística de los tests empleados en el análisis del DG es función de la intensidad del DG bajo la hipótesis alternativa de equilibrio gamético, así como de las frecuencias alélicas, el signo de las asociaciones y el tamaño muestral (BROWN 1975; WEIR y COCKERHAM 1978; CHAKRABORTY 1984; FU y ARNOLD 1992; ZAPATA y ALVAREZ 1992, 1993). De acuerdo con esto, la potencia puede ser muy baja si el DG no es intenso y/o las frecuencias alélicas no son intermedias. Bajo estas condiciones, los tamaños de muestra necesarios para detectar DG son muy elevados. Además, las asociaciones estadísticas entre loci en tablas de contingencia habían sido testadas tradicionalmente mediante el test exacto de Fisher, el cual es muy conservador (UPTON 1982; D'AGOSTINO *et al.* 1988; ZAPATA y ALVAREZ 1997). Esto aumenta en gran medida la probabilidad del error de tipo II (aceptar la hipótesis nula de equilibrio gamético cuando es falsa) y por tanto no es

deseable en estudios de DG. Así, el test chi-cuadrado es más apropiado para aplicar a una tabla de contingencia 2 x 2 y tiene la ventaja de que está relacionado con coeficientes que miden la intensidad del DG (HEDRICK *et al.* 1978; ZAPATA y ALVAREZ 1997). ZERBA *et al.* (1991), analizando estudios publicados de DG en la región del gen *Apo B*, informaron acerca de un elevado error de tipo II en estos trabajos. El reanálisis de varios estudios de DG conjuntamente, llevó a la observación de que el DG resultaba ser relativamente frecuente incluso entre pares de loci débilmente ligados (ZAPATA y ALVAREZ 1992). Parecía por tanto que el DG sí tenía cierta relevancia en las poblaciones naturales.

La estimación de la intensidad del DG, además de su detección, también es un punto problemático en muchos trabajos. En algunas ocasiones se han utilizado las probabilidades resultantes de los tests de significación como medida de la intensidad del DG. Pero esta medida no es una buena herramienta para medir la intensidad o hacer comparaciones entre loci porque la potencia estadística depende del tamaño de muestra, de las frecuencias alélicas y del signo de las asociaciones (BROWN 1975; THOMPSON *et al.* 1988; ZAPATA *et al.* 2001b). Así, la mejor manera de medir la intensidad del DG son los coeficientes o índices de asociación (KENDALL y STUART 1979; AGRESTI 1990; EVERITT 1997). Existen en la literatura múltiples coeficientes que miden la asociación entre pares de loci (HEDRICK 1987; DEVLIN y RISCH 1995; GUO 1997). Los más utilizados son el coeficiente D' (LEWONTIN 1964) y el r^2 (HILL y ROBERTSON 1968). Aunque no existe ningún coeficiente totalmente independiente de las frecuencias alélicas, el coeficiente D' presenta un rango de -1 a +1, independientemente de las frecuencias alélicas (ZAPATA 2000). Además de que sus propiedades se conocen bien para sistemas multialélicos (ZAPATA *et al.* 2001 a, b).

El DG en la detección de genes implicados en enfermedades

El clonaje posicional persigue asignar una posición en el genoma a un gen implicado en una enfermedad sin necesidad de un conocimiento previo de la biología de la enfermedad. Tras la identificación del gen responsable de la

enfermedad se puede llegar a un conocimiento de las bases moleculares y fisiológicas de la enfermedad, lo que sería un primer paso para el desarrollo de terapias adecuadas. Tradicionalmente, el clonaje posicional de un gen comienza con un análisis de ligamiento en familias que están segregando para el carácter de la enfermedad. En estas familias se estudia la cosegregación de la enfermedad con marcadores genéticos seleccionados con localización conocida en el genoma (LANDER y SCHORK 1994; RISCH 2000; BOTSTEIN y RISCH 2003). La correlación del patrón de herencia de la enfermedad y de los marcadores genéticos en una familia será tanto mayor cuanto menor sea la frecuencia de recombinación entre ambos. Así, los marcadores más próximos a la mutación de la enfermedad mostrarán una segregación conjunta con la enfermedad en un árbol familiar. El número de eventos recombinacionales en una familia es limitado, por lo que no es posible determinar el orden de los genes a cortas distancias, donde la recombinación es menos frecuente. Por esta razón, los análisis de ligamiento adscriben al gen implicado en la enfermedad a una región de una a varias megabases donde pueden estar presentes multitud de genes, haciendo el clonaje del gen responsable una tarea complicada. En estos casos, para disminuir el tamaño de la región del genoma posiblemente implicada en la enfermedad, se han obtenido resultados satisfactorios mediante la utilización del DG. Los patrones de DG en una población actual reflejan la acción de la recombinación a través de las generaciones pasadas, la historia evolutiva de la población, de manera que aumenta la muestra de recombinantes respecto a una sola familia (JORDE 1995). Así, al incorporar gran cantidad de eventos recombinacionales, los análisis de DG pueden aportar información acerca del ordenamiento de los loci en regiones que son demasiado pequeñas para mostrar recombinaciones en unas pocas familias. En un primer momento, la mutación de la enfermedad surge en DG completo en un haplotipo *background* de marcadores ligados. A través de las generaciones siguientes, la recombinación disipa gradualmente el DG entre la mutación de la enfermedad y los marcadores polimórficos. En ausencia de otros fenómenos poblacionales, la tasa de disipación del DG dependerá únicamente de la frecuencia de recombinación. Así, los marcadores más próximos a la mutación de la enfermedad (con una menor frecuencia de recombinación), manifiestan mayor

DG que los más alejados (JORDE 1995; TERWILLIGER 1995; TERWILLIGER y GÖRING 2000; BOTSTEIN y RISCH 2003).

La utilización del DG para la localización de factores genéticos implicados en enfermedades, depende de las asociaciones existentes entre los marcadores genéticos y la enfermedad. Estos análisis de asociación se basan en la comparación de las frecuencias de un marcador genético particular entre individuos afectados por la enfermedad, no relacionados, e individuos control, aleatorios, no afectados (BOTSTEIN y RISCH 2003). La asociación se dice que está presente cuando hay diferencias significativas en estas frecuencias entre las muestras de enfermos y controles (LANDER y SCHORK 1994; RISCH y MERIKANGAS 1996; OTT 1999; TERWILLIGER y GÖRING 2000). Si el marcador es anónimo, es decir, no tiene una relación de causalidad con la enfermedad, entonces se asocia con ella sólo si la mutación responsable del fenotipo enfermo está en las proximidades; marcador y mutación causante de la enfermedad estarían así en DG (SCHAID y SOMMER 1993; SPIELMAN y EWENS 1996).

Mediante análisis de ligamiento y posterior aplicación del DG para disminuir el tamaño de la región candidata, se han obtenido resultados espectaculares para multitud de genes implicados en enfermedades con herencia mendeliana, cuya base genética está en el defecto de un único gen, por lo menos dentro del contexto de una familia. Entre algunas de estas enfermedades se encuentran la fibrosis quística (KEREM *et al.* 1989; 1990) y la enfermedad de Huntington (GUSELLA *et al.* 1989; WHITE *et al.* 1997). Para estas enfermedades monogénicas, raras, hay una correspondencia casi total entre el fenotipo y el genotipo dentro de los árboles familiares. Por esto el mapeo de enfermedades monogénicas es más sencillo, en principio, que el de enfermedades complejas en las que el efecto de un gen puede aparecer enmascarado por factores medioambientales o por la presencia de otros genes que también influyen en el carácter (TERWILLIGER y GÖRING 2000; THOMSON 2001). Por otro lado, la realización de un análisis de ligamiento en familias requiere asumir un modelo genético preciso. Se debe especificar el modo de herencia, las frecuencias alélicas y la penetrancia de cada

genotipo. Si se asume que la penetrancia es baja, puede haber una disminución en la potencia para detectar el ligamiento. Este problema se agrava en el caso de una enfermedad compleja puesto que, en la mayoría de los casos, no es posible asumir ningún modelo de herencia específico (CARDON y BELL 2001; STRACHAN y READ 2004). El término de “enfermedades complejas” se refiere a cualquier enfermedad que no exhibe una herencia dominante o recesiva mendeliana clásica atribuible a un único locus, pudiendo incluirse los loci de penetrancia incompleta, fenocopias y herencia poligénica (LANDER y SCHORK 1994; KRUGLYAK y LANDER 1995). Estas enfermedades complejas están determinadas por un número variable de factores genéticos y ambientales de mayor o menor efecto (SCHORK 1997). Así, para detectar pequeños efectos de genes individuales, se necesita, o suficiente potencia estadística, o alguna forma de evaluar los loci colectivamente (LONGMATE 2001). RISCH y MERIKANGAS (1996) consideraron la estrategia de asociación en un sistema de un locus con dos alelos y concluyeron que los tests de asociación tienen suficiente potencia para detectar efectos modestos de loci individuales en desórdenes complejos.

Uno de los obstáculos con los estudios de asociación es la gran cantidad de falsos positivos que se producen. Existen en la literatura gran cantidad de estudios de asociación que no han podido ser corroborados (TERWILLIGER y WEISS 1998; WEISS y TERWILLIGER 2000; CARDON y BELL 2001; IOANNIDIS *et al.* 2001, 2003). Entre las razones para esta falta de reproducibilidad se encuentra una inadecuada selección del grupo control. La selección de los controles es crucial porque cualquier diferencia sistemática en las frecuencias alélicas entre los enfermos y los controles, puede aparecer como asociación con la enfermedad. Cuando existen diferentes frecuencias alélicas entre el grupo control y los enfermos, el resultado del análisis de asociación puede ser significativo, pero no relacionado con la presencia del alelo de la enfermedad en esa región del genoma. De esta manera, la asociación detectada sería espúrea dado que se debería a la estructuración poblacional subyacente (CHAKRABORTY y WEISS 1988; PRITCHARD y ROSENBERG 1999; CARDON y BELL 2001; ARDLIE *et al.* 2002; LOHMEUILLER *et al.* 2003). Así, es necesario que las frecuencias genotípicas observadas en las muestras control y

enfermos no difieran significativamente de las proporciones esperadas bajo equilibrio Hardy-Weinberg. De esta forma se podría descartar la existencia de fenómenos de subestructuración en la población de estudio.

El DG dependiente de marcador

Para realizar un correcto diseño de un estudio de asociación, y una correcta interpretación de los resultados, es importante conocer el DG de *background*, es decir, disponer de información previa acerca de la distribución del DG a lo largo de los cromosomas entre marcadores genéticos, de manera que no exista confusión con el DG entre la enfermedad y el locus marcador (FREIMER *et al.* 1997; TERWILLIGER y WEISS 1998; SERVICE *et al.* 2001; ZAPATA *et al.* 2001b). En el pasado, para la caracterización del DG, se utilizaron marcadores genéticos con dinámicas evolutivas muy diferentes. Pero los resultados obtenidos con marcadores distintos podrían no ser comparables, puesto que la cuantificación del DG podría verse distorsionada por una aportación de las características propias del marcador a los niveles de DG de *background*, es decir, debido a la existencia de DG dependiente de marcador. Entre los marcadores genéticos se encuentran los loci proteicos, utilizados durante la década de los 70 (SMOUSE y NEEL 1977; SINNOCK y SING 1992). Ya en la década de los 80 se utilizaron los RFLPs (*Restriction Fragment Length Polymorphism*) que utilizaban enzimas de restricción para detectar la presencia o ausencia de un sitio de restricción polimórfico (BOTSTEIN *et al.* 1980). Más recientemente, en la década de los 90, los RFLPs fueron sustituidos por los microsatélites (repeticiones nucleotídicas en tándem de 2 a 6 pb), como el marcador de elección en estudios poblacionales (SHRIVER *et al.* 1997; DI RIENZO *et al.* 1998; KIMMEL *et al.* 1998; ZHIVOTOVSKY *et al.* 2003), análisis de ligamiento y mapeo genético (DIB *et al.* 1996) y estudios de DG (PETERSON *et al.* 1995; ZAPATA *et al.* 2001b). Los microsatélites están presentes tanto en organismos eucariotas como procariotas, aunque su abundancia varía en diferentes especies (PRIMMER y ELLEGREN 1998) y tiende a estar correlacionada positivamente con el tamaño del genoma (HANCOCK 1996; TÓTH *et al.* 2000; MORGANTE *et al.* 2002). Además, muestran altos niveles de polimorfismo en el

número de sus unidades de repetición, están ampliamente distribuidos por el genoma y se pueden genotipar mediante amplificación por PCR (WEBER y MAY 1989). El análisis inicial de la secuencia del genoma humano concluyó que los microsatélites representan el 3 % de la secuencia total del genoma. Las repeticiones de tipo dinucleotídico dominan, seguidas por las mono-, tetra- y trinucleotídicas. Entre los dinucleótidos, las repeticiones (CA)_n son las más frecuentes (ELLEGREN 2004). Este tipo de dinucleótidos son especialmente útiles para análisis de ligamiento y asociación, pues se calcula que hay entre 50.000 y 100.000 repartidos a lo largo de todo el genoma humano (VALDES *et al.* 1993), con un promedio de uno cada 30 – 60 kb (RAMEL 1997; SCHLÖTTERER 1998). La gran mayoría de los microsatélites se encuentran en secuencias de DNA no codificadoras, tanto en regiones intergénicas como en intrones y han sido considerados como DNA “basura” o DNA “egoista” (NOWAK 1994). Aunque también se ha sugerido que podrían tener alguna función en la regulación de la expresión génica (KASHI *et al.* 1997; AKAI *et al.* 1999).

Los microsatélites se encuentran entre los polimorfismos de DNA con mayor variabilidad en el genoma (WEBER 1990). Esta variabilidad está caracterizada por una elevada heterozigosidad y por la presencia de un gran número de alelos, lo que contrasta fuertemente con el DNA de copia única. Los niveles de heterozigosidad en los microsatélites de tipo dinucleotídico de humanos han sido caracterizados en individuos de origen caucasoide en torno al 70 % (DEKA *et al.* 1995; DIB *et al.* 1996). Debido a estos niveles de heterozigosidad y al elevado número de alelos, estos marcadores proporcionan herramientas ideales para realizar análisis de ligamiento basados en pedigríes. Así, su aplicación ha llevado a la identificación de genes implicados en muchas enfermedades monogénicas y algunas poligénicas (COLLINS 1995). Además estos polimorfismos proporcionan suficiente potencia estadística para detectar DG a grandes distancias a lo largo del genoma (PETERSON *et al.* 1995; LAAN y PÄÄBO 1997; CHAPMAN y WIJSMAN 1998; ZAPATA *et al.* 2001b; OHASHI y TOKUNAGA 2003). Y han sido ampliamente utilizados para la construcción de mapas genéticos (DIETRICH *et al.* 1996; DIB *et al.* 1996), análisis forenses y estudios poblacionales (ELLEGREN 2004).

Su elevado polimorfismo es debido a la elevada tasa mutacional, que en humanos oscila entre 10^{-5} y 10^{-3} en cada generación (CHAKRABORTY *et al.* 1997; WEBER y WONG 1993; PRIMMER *et al.* 1998; BRINKMAN *et al.* 1998; XU *et al.* 2000). Además, su localización en regiones no codificadoras sugiere un comportamiento neutral en el proceso selectivo. De manera que su frecuencia y distribución debería reflejar por tanto el proceso mutacional subyacente. Así, el mecanismo mutacional que genera nuevos alelos parece ser el “*slippage*” o deslizamiento de cadena de la DNA polimerasa durante la replicación, que consiste en la disociación transitoria de las cadenas de DNA durante la replicación, seguida de una reasociación con un emparejamiento de bases erróneo (LEVINSON y GUTMAN 1987; SCHLÖTTERER y TAUTZ 1992; SCHLÖTTERER 2000).

Se han propuesto varios modelos de la dinámica evolutiva de los microsatélites que derivan, en su mayoría, del modelo de mutación por pasos o “stepwise mutation model” (OHTA y KIMURA 1973). Este modelo postula que cada mutación altera el tamaño de un alelo en la adición o supresión de una unidad de repetición (SHRIVER *et al.* 1993; VALDES *et al.* 1993; KIMMEL y CHAKRABORTY 1996; KIMMEL *et al.* 1996). DI RIENZO *et al.* (1994), tras examinar datos poblacionales en humanos, postularon un modelo de dos fases en el cual, además de los cambios en una sola unidad de repetición, existían mutaciones que modificaban el tamaño de un alelo en varias unidades de repetición. Según estos modelos no existirían constricciones en el tamaño de los alelos, de manera que éste tendería a aumentar indefinidamente. Sin embargo, existen numerosas evidencias de la existencia de un límite máximo en el tamaño de los microsatélites (GARZA *et al.* 1995). Así, se propusieron numerosos modelos que introducían un límite máximo en el tamaño de los alelos (NAUTA y WEISSING 1996; FELDMAN *et al.* 1997; POLLOCK *et al.* 1998) o diferencias en las tasas de mutación de forma que los alelos de mayor tamaño mutaran preferentemente a alelos de menor tamaño (GARZA *et al.* 1995; ZHIVOTOVSKY 1999; HUANG *et al.* 2002; CALABRESE y DURRETT 2003). Pero la evidencia experimental muestra que en el proceso mutacional aparece una preferencia hacia una ganancia de unidades de repetición, además de que los

alelos de mayor tamaño mutan más frecuentemente que los de menor número de unidades de repetición (WIERDL *et al.* 1997; BRINKMAN *et al.* 1998; FARRALL y WEEKS 1998). Esto no quiere decir que el tamaño de los microsatélites aumente indefinidamente. De hecho, éste podría estar gobernado por dos procesos operando simultáneamente. Mutaciones de ganancia o pérdida de unidades de repetición, junto a mutaciones puntuales que rompen la continuidad de los alelos, lo que disminuye la tasa de mutación (BELL y JURKA 1997; KRUGLYAK *et al.* 1998; CALABRESE *et al.* 2001), evitando de esa manera que el microsatélite se expanda indefinidamente.

La precisión de los análisis basados en microsatélites depende en gran medida del éxito alcanzado en minimizar los errores de genotipado. Estos errores pueden ser de naturaleza diversa, desde la incorrecta determinación del tamaño de un alelo, hasta el genotipado de un individuo heterocigoto como homocigoto debido a una no amplificación total o parcial de uno de los alelos. Las repeticiones dinucleotídicas son generalmente más difíciles de genotipar que las tri- y tetranucleotídicas (SCHLÖTTERER 1998). Los errores de genotipado en los dinucleótidos pueden surgir por varias razones incluyendo errores de procesamiento durante la reacción de PCR (ANDERSON *et al.* 2000) y el uso de condiciones electroforéticas inadecuadas para resolver los productos de amplificación, que podrían dar como resultado una falta de resolución o una escasa precisión en la asignación de los tamaños alélicos. Los métodos automáticos de análisis de microsatélites tratan de eliminar o minimizar las bandas llamadas “artefactuales” para una interpretación más sencilla del genotipo individual (SCHLÖTTERER 1998). En este sentido, recientemente se ha desarrollado un sistema electroforético en geles no desnaturizantes de utilidad para detectar esos errores de genotipado. Este sistema aprovecha el patrón de las bandas minoritarias, las mal llamadas “artefactuales”, de la electroforesis en condiciones nativas como control de un correcto genotipado (RODRIGUEZ *et al.* 2001; RODRIGUEZ y ZAPATA 2002).

Al final de los años 90 se asistió al resurgimiento de los SNPs (*Single Nucleotide Polymorphisms*), es decir, polimorfismos en un único nucleótido de la secuencia del DNA (CHAKRAVARTI 1998; WANG *et al.* 1998; KRUGLYAK y NICKERSON 2001). Durante los años 80 los SNPs se habían detectado mediante enzimas de restricción para identificar la presencia o ausencia de dianas de restricción, es decir, eran detectados en forma de RFLPs. Los SNPs son el tipo de variación genética más común en humanos y esta es la razón principal de su extensiva utilización actual como marcadores en estudios genéticos de variabilidad poblacional en humanos (CHAKRAVARTY 1998; 1999), selección natural (AKEY *et al.* 2002) y estudios de ligamiento en enfermedades (LANDER 1996; KRUGLYAK 1997). Sobre todo en los últimos años su principal aplicación está en estudios de DG de *background* (TAILLON-MILLER *et al.* 2000; WALL y PRITCHARD 2003; KE *et al.* 2004) y en la localización de genes implicados en enfermedades complejas (RISCH y MERIKANGAS 1996; KRUGLYAK 1999; MARTIN *et al.* 2000; PATTERSON *et al.* 2004; MANIATIS *et al.* 2004) y caracteres farmacogenéticos (HOUSMAN y LEDLEY 1998; GOLDSTEIN 2003). Aunque en principio en cada posición nucleotídica cualquiera de los cuatro nucleótidos puede estar presente, en la práctica, los SNPs son bialélicos (BROOKES 1999). La razón de este bialelismo hay que buscarla en la baja tasa de sustituciones nucleotídicas que presentan (en torno a 1×10^{-9} por generación y par de bases en mamíferos) (LI *et al.* 1981; LI y SADLER 1991; MARTINEZ-ARIAS *et al.* 2001). Por lo tanto, la probabilidad de que ocurran dos cambios independientes en la misma posición, es muy baja. Además, existe un sesgo en las mutaciones que resulta en la prevalencia de dos tipos de SNPs. Así, existe una tasa mutacional dos veces superior de transiciones (purina $\leftrightarrow$ purina, pirimidina $\leftrightarrow$ pirimidina) que de transversiones (purina $\leftrightarrow$ pirimidina) (WANG *et al.* 1998; VIGNAL *et al.* 2002). En humanos se observan todas las combinaciones de sustituciones polimórficas, siendo las sustituciones A/G (o su complementario T/C) las más frecuentes (WANG *et al.* 1998; MILLER y KWOK 2001).

Los SNPs se encuentran ampliamente distribuidos por todo el genoma. Estimaciones recientes sugieren una cifra de 10 millones de estos marcadores en el genoma humano (BOTSTEIN y RISCH 2003; FORTON *et al.* 2005), de manera que

podríamos encontrar uno cada 300 pb. Aunque existe una enorme diversidad en su frecuencia entre diferentes genes, lo que podría reflejar diferentes presiones selectivas en cada gen, así como diferentes tasas mutacionales y recombinacionales a lo largo del genoma. Así, se ha observado que la abundancia de SNPs en regiones codificadoras y no codificadoras es esencialmente la misma, aunque en regiones codificadoras la incidencia de SNPs que provocan cambios no sinónimos está reducida respecto a SNPs silentes (CHARLESWORTH *et al.* 1993; NACHMAN *et al.* 1998; CARGILL *et al.* 1999; HALUSHKA *et al.* 1999; GRAY *et al.* 2000).

Los SNPs también difieren entre sí en sus frecuencias alélicas. La frecuencia de un alelo está relacionada con su edad en la población. Según la teoría genético-poblacional, para cualquier locus polimórfico considerado neutral, la probabilidad de que un alelo represente el alelo más antiguo en la población es igual a su frecuencia (WATTERSON y GUESS 1977). El hecho de que los SNPs raros sean jóvenes o viejos en la población humana tiene profundas implicaciones en su utilidad en los estudios de asociación enfermedad-marcador (CHAKRAVARTY 1999). Un SNP con un alelo joven mantendrá un segmento mayor en DG a su alrededor que un SNP con un alelo antiguo que habrá tenido más tiempo para experimentar una recombinación considerable con otros haplotipos (CHAKRAVARTI 1998). Así, MOFFATT *et al.* (2000) obtuvieron un efecto significativo de la frecuencia alélica en la extensión del DG, de manera que las combinaciones de alelos raros mostraban mayor DG que las combinaciones de alelos comunes. También la frecuencia alélica está relacionada con la diferenciación poblacional de manera que los SNPs relativamente comunes en la población humana, son compartidos generalmente por la mayoría de las poblaciones y deben haber surgido antes de la expansión de las poblaciones actuales (BROOKES 1999; HALUSHKA *et al.* 1999; KRUGLYAK 1999).

Las características óptimas de los marcadores genéticos para su utilización en estudios de DG ha sido un tema planteado en numerosas ocasiones (SLATKIN 1994; OTT y RABINOWITZ 1997; CHAPMAN y WIJSMAN 1998; XIONG y JIN 1999; SHAM *et al.* 2000; OHASHI y TOKUNAGA 2003). Disponibilidad de un elevado número de marcadores, facilidad y alta capacidad de las técnicas de genotipado, y suficiente

potencia estadística para detectar DG, son características muy deseables a la hora de elegir el marcador genético más apropiado. En lo que respecta a los microsatélites, se dispone de un número bastante elevado, localizados por todo el genoma (entre 50.000 y 100.000) y su sistema de genotipado es a priori sencillo mediante amplificación por PCR. En cambio los SNPs son mucho más numerosos y en los últimos años han surgido gran cantidad de consorcios encargados de la caracterización de nuevos SNPs en diferentes poblaciones (TSC, *The SNP Consortium*, <http://snp.cshl.org>), de manera que podríamos disponer de un SNP cada 300 pb (FORTON *et al.* 2005). Las técnicas actuales de genotipado tienen una elevada capacidad para procesar gran cantidad de SNPs en poco tiempo. Además, la interpretación de los genotipos para los SNPs es más sencilla que para loci microsatélites, dado que sólo presentan dos alelos y la posibilidad de cometer un error de genotipado es menor (VIGNAL *et al.* 2002).

Estos marcadores presentan diferencias en su potencia estadística para detectar el DG. A este respecto, OTT y RABINOWITZ (1997) abordaron el estudio del efecto de la heterozigosidad del marcador en la potencia para detectar DG. Estos autores concluyeron que la potencia para detectar DG aumenta con una mayor heterozigosidad de los marcadores implicados. En la misma línea, CHAPMAN y WIJSMAN (1998), XIONG y JIN (1999) y SHAM *et al.* (2000), concluyeron que marcadores altamente polimórficos (microsatélites) tienen mayor potencia que marcadores bialélicos (SNPs) en la detección del DG. Para alcanzar la misma potencia con marcadores bialélicos que con multialélicos se requeriría un tamaño de muestra mucho mayor y un espaciado entre marcadores menor que en el mismo estudio con marcadores multialélicos (CHAPMAN y WIJSMAN 1998). Además, los microsatélites de elevada heterozigosidad pueden ser informativos para localizar variantes raras de enfermedades, mientras que los SNPs comunes (utilizados en la mayoría de los trabajos) sólo serían capaces de localizar variantes comunes (PRITCHARD 2001; PRITCHARD y COX 2002; VARILLO *et al.* 2003). Hasta el momento no existe un acuerdo claro acerca del tipo de marcador más adecuado para realizar estudios de DG. Así, los marcadores microsatélites parecen ser más informativos que los SNPs para la detección del DG, sin embargo, la mayor

cantidad de SNPs y su, a priori, menor tasa de errores en el genotipado implicarían que sería posible obtener un patrón de DG preciso utilizando SNPs a una mayor densidad en el genoma (OHASHI y TOKUNAGA 2003).

Actualmente la literatura está repleta de estudios que caracterizan la extensión e intensidad del DG de *background* a largas y cortas distancias utilizando marcadores genéticos con dinámicas evolutivas muy diferentes, como son los **microsatélites** (PETERSON *et al.* 1995; CROUAEU-ROY *et al.* 1996; TISHKOFF *et al.* 1996; LAAN y PÄÄBO 1997; TISHKOFF *et al.* 1998; HUTTLEY *et al.* 1999; EAVES *et al.* 2000; GORDON *et al.* 2000; TISHKOFF *et al.* 2000; WILSON y GOLDSTEIN 2000; MOHLKE *et al.* 2001; SERVICE *et al.* 2001; ZAPATA *et al.* 2001b; HALL *et al.* 2002; RYBICKI *et al.* 2002; AULCHENKO *et al.* 2004) y los **SNPs** (BONNEN *et al.* 2000; DUNNING *et al.* 2000; EISENBARTH *et al.* 2000; MARTIN *et al.* 2000; MOFFATT *et al.* 2000; TAILLON-MILLER *et al.* 2000; EISENBARTH *et al.* 2001; REICH *et al.* 2001; TIRET *et al.* 2002; CLARK *et al.* 2003; SHIFMAN *et al.* 2003; CARLSON *et al.* 2004; KE *et al.* 2004; TSUNODA *et al.* 2004). También se han realizado trabajos combinando datos de microsatélites y datos de SNPs (ZAVATTARI *et al.* 2000; ABECASIS *et al.* 2001; KAESMANN *et al.* 2002; VARILLO *et al.* 2003). Sin embargo, en estos trabajos no se consideró la posibilidad de que una parte del DG de *background* detectado, fuese debido a las características particulares de cada tipo de marcador. Además, este DG dependiente de marcador podría ser importante para el diseño e interpretación de los estudios de localización de genes de enfermedades mediante DG.

La identificación de genes implicados en enfermedades comunes de carácter complejo en humanos, depende del DG existente entre el carácter y los marcadores analizados. Así, el número de marcadores necesarios para realizar un estudio de asociación a lo largo de todo el genoma, depende de la distancia hasta la que se puede detectar el DG. De los estudios de DG entre loci microsatélites se puede extraer la conclusión de que éste se extiende a grandes distancias, típicamente de varias megabases (EAVES *et al.* 2000; PRITCHARD y PRZEWORSKI 2001; KAUPPI *et al.* 2003). PETERSON *et al.* (1995) fueron los primeros en estudiar el DG de *background* con loci microsatélites a lo largo de una región extensa,

examinando 32 microsatélites en varias regiones del cromosoma 4 con una muestra de población finlandesa y obtuvieron DG a lo largo de distancias superiores a 1 cM o 1 Mb. Muchos otros trabajos con loci microsatélites obtuvieron DG a distancias superiores de 4 cM (EAVES *et al.* 2000; GORDON *et al.* 2000; WILSON y GOLDSTEIN 2000; MOHLKE *et al.* 2001; SERVICE *et al.* 2001; HALL *et al.* 2002; KAESMANN *et al.* 2002; AULCHENKO *et al.* 2004). ZAPATA *et al.* (2001b) obtuvieron DG a lo largo de la región telomérica del cromosoma 11, entre loci microsatélites separados por distancias hasta de 19 cM o 14 Mb, con un promedio de 7 cM. La mayor cantidad de asociaciones significativas observadas entre microsatélites separados por grandes distancias, respecto a estudios previos, se debía a la utilización de estimas más fiables de los niveles de DG a lo largo de una región, que, además, ofrecían una mayor potencia estadística. Estos autores estimaron la intensidad y significación del DG considerando exclusivamente las asociaciones de signo positivo, debido a su menor varianza de muestreo y mayor potencia estadística para ser detectadas por un test de significación (BROWN 1975; ZAPATA *et al.* 2001a, b). En este trabajo, dado que detectaron DG incluso entre los microsatélites analizados más alejados entre sí, no se estableció un límite máximo para la extensión del DG entre pares de microsatélites. Por lo tanto, se hace necesario un estudio empleando la misma metodología de estimación del DG, en una región más extensa que la analizada por ZAPATA *et al.* (2001b), para tratar de averiguar la distancia máxima a la que es posible detectar DG o si esos patrones de DG se extienden por todo el brazo cromosómico.

ZAPATA *et al.* (2001b) observaron, además, que los patrones de DG dependían de los alelos implicados. Mediante el análisis del DG interalélico positivo entre pares de alelos de los microsatélites, observaron que los alelos más raros y de tamaño extremo tendían a presentar DG más frecuentemente que los alelos restantes. Este DG dependiente de alelo tiene implicaciones para la localización de genes implicados en enfermedades. Estudios posteriores deberán corroborar estos patrones dependientes de alelo para regiones cromosómicas más extensas, y diferentes poblaciones.

Los estudios de DG de *background* realizados entre pares de SNPs no encuentran asociaciones significativas a tan largas distancias como ocurría entre pares de microsatélites. KRUGLYAK (1999) predijo mediante un estudio de simulación que el DG útil para realizar un estudio de mapeo en todo el genoma con SNPs no se extendería más allá de 3 kb en la población general. De esta manera harían falta del orden de 500.000 SNPs para un estudio de asociación a nivel de todo el genoma. Pero simulaciones posteriores (COLLINS *et al.* 1999) y observaciones empíricas han demostrado que esta observación era demasiado pesimista (BOEHNKE 2000; OTT 2000; WEISS y TERWILLIGER 2000). Así, DUNNING *et al.* (2000) encontraron que los pares de SNPs separados hasta 20 kb mostraban DG significativo. ABECASIS *et al.* (2001), examinando el DG en tres regiones cromosómicas en población del Reino Unido y nórdicos europeos, obtuvieron un 50 % de SNPs en DG separados por distancias de hasta 50 kb y también en algunos pares de SNPs separados por distancias de 500 kb. Estos datos son consistentes con otros trabajos donde se observa que el DG entre pares de SNPs se extiende a distancias considerables, como varios cientos de kilobases (CLARK *et al.* 1998; COLLINS *et al.* 1999; RIEDER *et al.* 1999; MOFFATT *et al.* 2000; TAILLON-MILLER *et al.* 2000; TEMPLETON *et al.* 2000; REICH *et al.* 2001; TSUNODA *et al.* 2004).

Los diferentes resultados reflejados en la literatura entre SNPs y entre microsatélites, parecen indicar un comportamiento particular de cada uno frente al DG. Estos marcadores se utilizan muy frecuentemente en estudios de DG, pero se desconoce la influencia que podría ejercer las características propias de cada marcador en los niveles detectados. Así, la cuantificación de la cantidad de DG de *background* podría verse distorsionada debido a la existencia de DG dependiente de marcador. Son escasos los trabajos presentes en la literatura que estudien los patrones del DG entre pares SNPs y entre pares de microsatélites con la misma muestra y en la misma región cromosómica (SCHULZE *et al.* 2002; VARILLO *et al.* 2003; TANAKA *et al.* 2005). Estos trabajos informan de una mayor extensión de las asociaciones entre pares de microsatélites que entre pares de SNPs, probablemente debido a una mayor potencia estadística para detectar DG con loci multialélicos. Sin embargo no tuvieron en cuenta el efecto del marcador en los

niveles de DG obtenidos, de manera que no se conoce qué fracción de las asociaciones observadas son debidas a la dinámica mutacional particular de cada uno. Así, para separar el DG de *background* del DG dependiente de marcador, sería necesaria una cuantificación pormenorizada de los niveles de DG de *background* para una región cromosómica determinada, utilizando marcadores genéticos con dinámicas mutacionales muy diferentes, como son los SNPs y los microsatélites. Además, la metodología estadística empleada debe ser adecuada para detectar de manera fiable los niveles de DG con ambos tipos de marcadores genéticos.

La extensión del DG depende también de la localización cromosómica (HUTTLEY *et al.* 1999; TAILLON-MILLER *et al.* 2000; DALY *et al.* 2001; GABRIEL *et al.* 2002). Así, por ejemplo, HUTTLEY *et al.* (1999) estudiando los patrones de DG entre pares de microsatélites a lo largo de todo el genoma, concluyeron que el DG es dependiente de la región cromosómica analizada. Sus resultados evidenciaron DG a distancias de hasta 4 cM con una correlación negativa entre éste y la distancia genética, así como diferencias en la magnitud y distribución del DG, tanto dentro de cada cromosoma como entre ellos. Igualmente, TAILLON-MILLER *et al.* (2000) estudiando los patrones de DG entre pares de SNPs a lo largo de la región Xq25-Xq28, evidenciaron dos regiones de fuerte DG que se extendían a lo largo de 1 Mb de DNA en Xq25 y a lo largo de 340 kb en Xq28. Además, estudios empíricos han demostrado un mayor decaimiento del DG con la DF en regiones teloméricas respecto a regiones más centroméricas (WATKINS *et al.* 1994; TIRET *et al.* 2002).

El mantenimiento de una correlación negativa entre la distancia física y la intensidad del DG es un punto crítico para poder realizar estudios de DG para localizar genes de enfermedades. Así, según la teoría genético poblacional, el DG disminuye al aumentar la frecuencia de recombinación entre los pares de marcadores analizados, o entre el marcador y el gen implicado en la enfermedad que se trata de localizar (HILL y ROBERTSON 1968; JORDE 1995). De esta forma los marcadores más cercanos al gen implicado en la enfermedad manifestarán mayor

cantidad de asociaciones no al azar que aquellos más alejados (JORDE *et al.* 1994; JORDE 2000), lo que nos dará una idea de la localización de la mutación implicada en la enfermedad. Pero existe una variedad de fenómenos que pueden alterar esta relación negativa. Así, entre estos factores se incluye el tipo de marcador genético utilizado, junto con la edad de las mutaciones, mezcla de poblaciones, deriva genética, conversión génica, selección y desviaciones del apareamiento al azar (LITT y JORDE 1986; THOMPSON *et al.* 1988; JORDE *et al.* 1993, 1994; DUNNING *et al.* 2000). JORDE *et al.* (1994) obtuvieron una correlación negativa entre el DG y la distancia física para pares de marcadores separados por más de 60 kb. A menores distancias la variación del DG parecía estar menos relacionada con la frecuencia de recombinación, ya que factores como la mutación o la conversión génica podrían estar perturbando el patrón de DG. Tampoco SHIFMAN *et al.* (2003) obtuvieron correlación entre el DG y la DF a cortas distancias. Sin embargo otros trabajos, como el de DUNNING *et al.* (2000) y MOFFATT *et al.* (2000) entre otros, sí encontraron correlación a cortas distancias.

Los estudios teóricos han sugerido que el DG es altamente dependiente de la historia evolutiva poblacional (SLATKIN 1994), de manera que las poblaciones de rápido crecimiento mostrarían un menor DG de *background* respecto a las poblaciones de tamaño constante. Esta teoría fue comprobada experimentalmente mediante comparación de los patrones de DG resultantes con siete microsatélites del cromosoma X. Estos patrones se estudiaron en cuatro poblaciones y se encontró que la población de tamaño constante mostraba DG a mayores distancias (hasta 5,8 cM) respecto a las tres poblaciones de expansión rápida en las que apenas se encontraba DG incluso entre loci próximos (LAAN y PÄÄBO 1997).

La historia demográfica de una población podría predecir la utilidad relativa de una población para el mapeo por DG de loci implicados en enfermedades. Así, los alelos que llevan mutaciones de una enfermedad generadas durante una rápida expansión poblacional estarán rodeados por regiones de un elevado DG al haber tenido poco tiempo para que actuara la recombinación. Por el contrario,

estas poblaciones mostrarán bajos niveles de DG de *background* (SLATKIN 1994). Así, las poblaciones que han sufrido una expansión reciente (como la finlandesa) son más adecuadas para el mapeo de enfermedades monogénicas causadas por mutaciones recientes (DE LA CHAPELLE y WRIGHT 1998) que poblaciones de tamaño constante (como la población saami). En cambio, la situación parece ser al contrario en cuanto a las enfermedades complejas se refiere. Los alelos de susceptibilidad a estas enfermedades, generalmente comunes y antiguos en la población, estarán rodeados por áreas extensas de DG originado por deriva genética en poblaciones de tamaño constante. En cambio mostrarán muy poco DG en poblaciones de reciente expansión donde la acción de la deriva es insignificante (SLATKIN 1994; LAAN y PÄÄBO 1997; TERWILLIGER *et al.* 1998; NORDBORG y TAVARÉ 2002). Una ventaja indiscutible de las poblaciones aisladas de pequeño tamaño es que la enfermedad de estudio presentará una heterogeneidad genética menor, de manera que la localización de las variantes de susceptibilidad a la enfermedad existentes en esa población será más sencilla al reducirse el número de loci potencialmente implicados en la enfermedad (HÄSTBACKA *et al.* 1992; KRUGLYAK 1997; SHEFFIELD *et al.* 1998; WRIGHT *et al.* 1999; WEISS y TERWILLIGER 2000).

En general, el DG con loci microsatélites y con SNPs se extiende a mayores distancias en poblaciones aisladas de pequeño tamaño que se han mantenido estables durante largo tiempo (CROUAEU-ROY *et al.* 1996; LAAN y PÄÄBO 1997; GORDON *et al.* 2000; JORDE *et al.* 2000; WILSON y GOLDSTEIN 2000; ZAVATTARI *et al.* 2000; REICH *et al.* 2001; HALL *et al.* 2002; KAESSMANN *et al.* 2002; VARILLO y PELTONEN 2004). Aunque existen trabajos con microsatélites y con SNPs en los que no se observan diferencias en el patrón de DG entre poblaciones aisladas, en expansión, y poblaciones grandes con apareamiento aleatorio (EAVES *et al.* 2000; TAILLON-MILLER *et al.* 2000).

A pesar de la aparente complejidad de los patrones de DG observados en las diferentes poblaciones, estudios recientes han propuesto que la estructura subyacente del DG en el genoma humano se puede describir usando un marco de

trabajo relativamente sencillo en el cual el DG se distribuye a lo largo del genoma en bloques haplotípicos discretos de hasta 800 kb donde el DG decae muy lentamente con la distancia (DALY *et al.* 2001; PATIL *et al.* 2001; GOLDSTEIN 2001; DAWSON *et al.* 2002; GABRIEL *et al.* 2002; WALL y PRITCHARD 2003a, b). Se ha observado que los bloques contiguos están separados entre sí por puntos calientes de recombinación donde el DG decae rápidamente con la distancia (JEFFREYS *et al.* 2001; KAUPPI *et al.* 2003; PHILLIPS *et al.* 2003; STUMPF y GOLDSTEIN 2003). Además, algunos trabajos han sugerido que se debe tener en cuenta la historia poblacional, así como la heterogeneidad en las tasas de recombinación locales, para comprender el origen de los bloques haplotípicos (REICH *et al.* 2002; WANG *et al.* 2002; PHILLIPS *et al.* 2003; ZHANG *et al.* 2003; ANDERSON y SLATKIN 2004). De esta manera, los bloques haplotípicos podrían ser menos frecuentes, o de mayor tamaño debido a la fusión de bloques entre sí, en poblaciones que han sufrido profundos cuellos de botella en las cuales el DG se extiende más allá de los puntos calientes de recombinación (GABRIEL *et al.* 2002; STUMPF y GOLDSTEIN 2003; WALL y PRITCHARD 2003).

El modelo de bloques haplotípicos tiene importantes implicaciones para la localización de genes mediante estudios de asociación, ya que las fuertes asociaciones entre SNPs en cada bloque haplotípico sugieren la posibilidad de poder genotipar sólo unos pocos, que proporcionen información acerca del resto de SNPs en esa región o bloque haplotípico. Así, los principales haplotipos en cada bloque podrían etiquetarse con un pequeño número de SNPs, *haplotype-tagging* SNPs o htSNPs (JOHNSON *et al.* 2001), de manera que se reduciría el número total de SNPs necesario para realizar un estudio de asociación a lo largo de todo el genoma. Además, la extensión del DG varía dramáticamente a lo largo del genoma (DALY *et al.* 2001; JEFFREYS *et al.* 2001; REICH *et al.* 2001; GABRIEL *et al.* 2002; DAWSON *et al.* 2002) por lo que no sería eficiente el uso de SNPs seleccionados aleatoriamente o incluso regularmente espaciados, para realizar un estudio de asociación a lo largo de todo el genoma. En este sentido, se ha propuesto que la localización de genes con haplotipos es más potente que con SNPs individuales (RIOUX *et al.* 2001; STEPHENS *et al.* 2001).

Motivado por los estudios anteriores, se ha desarrollado en los últimos años el proyecto HapMap, formalmente iniciado en 2002 por grupos de investigación de cuatro continentes. El objetivo es caracterizar los patrones de DG entre SNPs a lo largo de todo el genoma en poblaciones de África, Asia y Europa, con el fin de facilitar los estudios de asociación para el descubrimiento de genes implicados en enfermedades complejas en humanos (SCHULZE *et al.* 2004; THE INTERNATIONAL HAPMAP CONSORTIUM 2003; ZAPATA *et al.* 2005).

EL CASO DE LA OTOSCLEROSIS

La otosclerosis clínica es un tipo de sordera de conducción que aparece entre la segunda y tercera década de vida, causada por una reabsorción ósea y reposición de tejido óseo anormal en la capa endocondral del hueso temporal (cápsula ótica) (MORRISON 1967; GORDON 1989; MENDER & TANGE 2003). El foco otosclerótico incipiente está constituido, sobre todo, por osteoclastos que destruyen el hueso laminar maduro. Posteriormente los osteoblastos realizan una reposición de hueso en esas áreas pero con un patrón celular y fibrilar anormal. El lugar predilecto para la aparición del foco otosclerótico es la región de la cápsula ótica, alrededor de la ventana oval, esto es, en la base del estribo (MENDER & TANGE 2003). El estribo es el último de los elementos que componen la cadena de huesecillos en el oído medio cuya función es conducir el sonido desde el tímpano hasta la ventana oval (véase la figura 1). Al formarse el foco otosclerótico se produce la fijación del estribo a la ventana oval y así no es capaz de conducir correctamente el sonido al oído interno desencadenándose una hipoacusia de conducción. La aparición de hipoacusia neurosensorial ocurre cuando la otosclerosis difunde al oído interno, y en este caso se habla de otosclerosis coclear (MIM166800). Pueden existir focos otoscleróticos en otras áreas que no comprometen a la audición y en este caso hablamos de otosclerosis histológica que sólo se pone de manifiesto en autopsias con una prevalencia del 8 – 10% (VAN DEN BOGAERT *et al.* 2001).

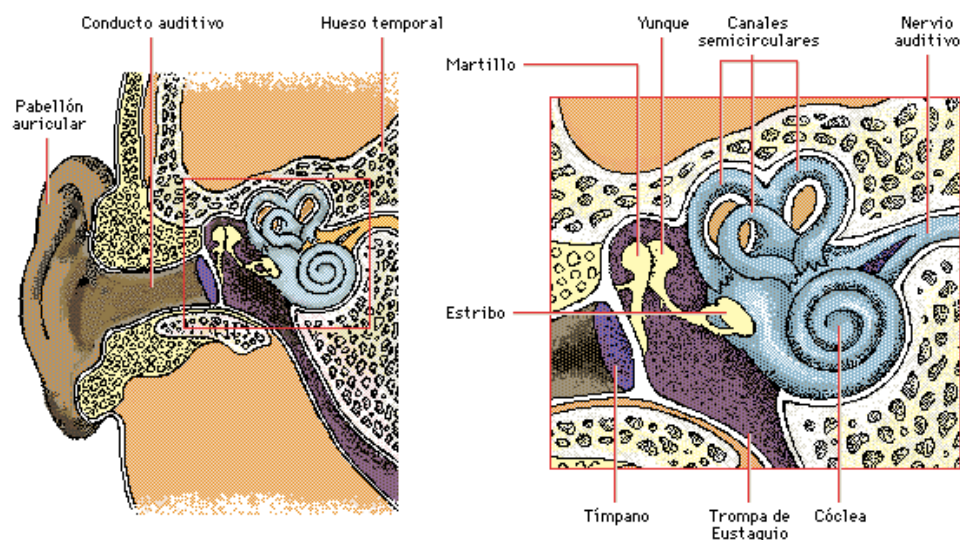


Figura 1. Esquema representativo del oído donde se señala la situación del estribo en la cadena de huesecillos, en el oído medio.

El cuadro sintomático comienza habitualmente a partir de los 20 años con una hipoacusia de conducción y, más rara vez neurosensorial, irregularmente progresiva, acelerada en ocasiones por circunstancias tales como embarazos o lactancias. La hipoacusia es bilateral en un 80% de los casos aproximadamente, aunque la pérdida suele ser mayor en el oído donde primeramente se manifestó la enfermedad (MORRISON 1967; GORDON 1989; MENDER & TANGE 2003).

La otosclerosis clínica se presenta en la población con una prevalencia de 0,2 – 1% en individuos de origen caucasoide (GORDON 1989) y una edad de aparición que oscila entre los 15 y 45 años (MORRISON 1967). Factores hormonales parecen influir en la aparición de la enfermedad ya que en la mayoría de los estudios la proporción de mujeres a hombres afectados es de 2:1 (LARSSON 1960; MORRISON 1967; GORDON 1989; BEN ARAB *et al.* 1993; MENDER & TANGE 2003). Aunque esta proporción parece desaparecer al estudiar los casos de otosclerosis histológica (ALTMAN *et al.* 1967; MENDER & TANGE 2003).

La etiología de la otosclerosis es compleja, probablemente debida a la interacción de factores genéticos y ambientales. Numerosos estudios apoyan la hipótesis de una herencia autosómica dominante con penetrancia incompleta (MORRISON 1967; GORDON 1989; BEN ARAB *et al.* 1993). En el pasado se han postulado una variedad de teorías diferentes para explicar el origen de la otosclerosis. De entre los factores hormonales propuestos como implicados en el desarrollo de esta enfermedad se encuentra la hormona paratiroidea (PTH) (GRAYELI *et al.* 1999). La función alterada de PTH provocaría niveles anormales de calcio y fósforo que estarían implicados en el recambio anormal de hueso observado durante la otosclerosis (GRAYELI *et al.* 1999). También la alteración en el nivel de algunas enzimas se ha relacionado con la etiología de la otosclerosis en varias ocasiones. Sin embargo, en algunos trabajos se ha sugerido que es el resultado de un proceso de remodelación ósea ya comenzado en el cual deben tener un efecto agonista (MENDER & TANGE 2003). Otros investigadores intentaron buscar relación entre la enfermedad y el grupo sanguíneo AB0, de manera que el factor implicado en la enfermedad estaría próximo en el cromosoma al grupo sanguíneo AB0 (MORRISON 1967). Respecto a la especificidad del complejo mayor de histocompatibilidad, SINGHAL *et al.* (1999) encontraron un aumento significativo de los antígenos HLA-A9 y HLA-A11 entre los hombres afectados por otosclerosis, comparado con un aumento del HLA-A3 en las mujeres afectadas. Otros autores habían encontrado, en pacientes que portaban los antígenos B8/DR3, una incidencia significativamente mayor en tres desórdenes de la audición, la enfermedad de Menier, otosclerosis y presbiacusia sensorial (BERNSTEIN *et al.* 1996).

Evidencias recientes sugieren que la otosclerosis es un desorden genéticamente complejo que implica regiones de susceptibilidad en los cromosomas 6, 7, 15 y 3 (TOMEK *et al.* 1998; CHEN *et al.* 2002; VAN DEN BOGAERT *et al.* 2001, 2002, 2004). TOMEK *et al.* (1998) realizaron un análisis de ligamiento en una familia del Sur de la India utilizando para ello 160 loci dispersos por todo el genoma. De esta manera identificaron una región de 14,5 cM en 15q25-q26. Así localizaron el primer locus implicado en la otosclerosis, denominado OTSC1 (MIM

166800). El locus OTSC1 todavía no está clonado, pero en este intervalo aparece el gen que codifica para el agregano que, después del colágeno, es el componente más numeroso de la matriz extracelular del cartílago (TOMEK *et al.* 1998). Sin embargo, VAN DEN BOGAERT *et al.* (2001) no encontraron ligamiento en la región del locus OTSC1. En cambio sí encontraron ligamiento estudiando una familia belga, en 16 cM de la región cromosómica 7q34-36, entre los marcadores D7S495 y D7S2426, pasando a ser el segundo locus identificado para la otosclerosis (OTSC2, MIM 605727). Esta región también incluye varios genes conocidos que podrían intervenir en la etiología de la otosclerosis (VAN DEN BOGAERT *et al.* 2001). El tercer locus identificado como implicado en la otosclerosis fue estudiado mediante análisis de ligamiento en una familia chipriota. En una región de 17,4 cM del cromosoma 6p21.3-22.3 se identificó el locus OTSC3 (MIM 608244) entre los marcadores GAAT3A06 y D6S1680 que incluye la región HLA, lo cual está de acuerdo con estudios previos en los que se relacionaba los antígenos HLA con la otosclerosis (CHEN *et al.* 2002; VAN DEN BOGAERT *et al.* 2002). Además, un cuarto locus, OTSC4, ha sido reservado por el comité de nomenclatura de la *Human Genome Organisation*, pero todavía no se ha publicado. El último locus identificado hasta la fecha se estudió en una familia holandesa de múltiples miembros segregando para la otosclerosis de forma autosómica dominante. Tras la exclusión de la implicación de los loci conocidos (OTSC1, OTSC2 y OTSC3), VAN DEN BOGAERT *et al.* (2004) localizaron mediante análisis de ligamiento un quinto locus, OTSC5, en un intervalo de 15,5 Mb de la región cromosómica 3q22-24 (MIM 608787).

Una hipótesis largamente mantenida y controvertida mantiene que la otosclerosis representa una manifestación local de un desorden general que implica errores en la estructura o remodelación de la matriz del tejido conectivo (GORDON *et al.* 1992; MCPHEE *et al.* 1993). En estos trabajos los autores encontraron una expresión anormal de enzimas que actúan en la remodelación del tejido conectivo. Pero no era probable que ésta fuese la causa de la patología sino sólo uno de sus síntomas, ya que el mecanismo regulador que gobierna la remodelación del tejido conectivo podría estar intentando disminuir su

transcripción para reducir la destrucción del tejido conectivo, generado por otra causa desconocida (MCPHEE *et al.* 1993).

- *Los genes del colágeno tipo I y la otosclerosis*

Más de cuatrocientas mutaciones en genes que codifican para tipos de colágeno diferentes causan una variedad de enfermedades que incluyen osteogénesis imperfecta (OI), condrodisplasias, algunas formas de osteoporosis, algunas formas de osteoartritis, y la enfermedad renal conocida con el nombre de síndrome de Alport (PROCKOP & KIVIRIKKO 1995). De entre todas estas mutaciones, 200 están en los genes COL1A1 y COL1A2 que codifican para las cadenas $\text{pro}\alpha 1(\text{I})$ y $\text{pro}\alpha 2(\text{I})$, respectivamente, del procolágeno tipo I (precursor del colágeno tipo I) (PROCKOP 1990). La mayoría de estas mutaciones han sido identificadas en pacientes con OI (SYKES *et al.* 1990; WILLING *et al.* 1992; WILLING *et al.* 1994; REDFORD-BADWAL *et al.* 1996) y estas diferentes mutaciones en los genes del colágeno tipo I dan lugar a distintos grados de severidad en la manifestación clínica. Un enfermo de OI se caracteriza por tener huesos frágiles y quebradizos. También otros tejidos ricos en colágeno tipo I se ven afectados de forma que aparece azulada la capa esclerótica de los ojos, dientes anormales, piel fina, tendones débiles y, en algunas formas una pérdida de oído similar a la ocurrida en la otosclerosis. En las formas más severas el afectado muere antes de nacer o poco después (PROCKOP y KIVIRIKKO 1995).

En un principio se pensó que la otosclerosis era una manifestación local de formas leves de OI debido a que comparten las mismas características clínicas e histopatológicas a nivel auditivo (OGILVIE y HALL 1962; MCKENNA *et al.* 1998, 2002; MENDER & TANGE 2003). Esta similitud con la otosclerosis, hace de los genes COL1A1 y COL1A2 unos buenos candidatos a factores implicados en la otosclerosis. Así, se puede realizar un análisis de asociación con marcadores genéticos localizados en estos dos genes, en lugar de utilizar la totalidad del genoma como marco de búsqueda. MCKENNA *et al.* (1998) analizaron estos dos genes y encontraron asociación significativa de tres marcadores polimórficos

dentro del gen COL1A1 con la otosclerosis clínica. Por el contrario no obtuvieron asociación con el gen COL1A2. Este hallazgo sugiere que el gen COL1A1 podría jugar un papel en la etiología de la otosclerosis. Además, en pacientes con formas muy leves de OI tipo I, ocurren varias mutaciones en COL1A1 pero no en COL1A2 (WILLING *et al.* 1994; 1996). Así, algunas formas leves de OI y la otosclerosis estarían relacionadas entre sí, ambas causadas por defectos en el gen COL1A1. En cambio autores como BROSANAN *et al.* (1977) y PEDERSON *et al.* (1985) encontraron diferencias significativas en el grado de desorganización del tejido conectivo entre la OI y la otosclerosis. Así, actualmente estas dos enfermedades se consideran como patologías independientes (MIM 166200).

La gran cantidad de factores genéticos que parecen estar implicados en la otosclerosis, podría variar en diferentes grupos poblacionales. Además, la observación de falta de reproducibilidad de estudios con otras muchas enfermedades (CARDON y BELL 2001), hace necesaria la realización de estudios que corroboren esos resultados, tanto en la misma población, como en otras.

OBJETIVOS

- I. El primer objetivo en este trabajo será la caracterización de los patrones de desequilibrio gamético (DG) de *background* dependientes de marcador, a lo largo de una extensa región anónima del genoma humano, como es el brazo corto del cromosoma 11 humano (11p). Esta caracterización se realizará mediante dos tipos de marcadores genéticos con dinámicas evolutivas bien diferenciadas, en una muestra de 405 individuos de la población gallega no relacionados entre sí. Así, se analizará el DG entre todos los pares resultantes de la combinación de 18 SNPs y entre todos los pares resultantes de la combinación de 22 microsatélites, todos ellos localizados a lo largo de 11p.

Los niveles de DG a lo largo de 11p, se evaluarán mediante la estimación de la frecuencia e intensidad del DG entre pares de SNPs y entre pares de microsatélites. El análisis del DG entre pares de SNPs se realizará independientemente del signo de las asociaciones. Por su parte, el DG entre pares de microsatélites se analizará sólo teniendo en cuenta las asociaciones de signo positivo, a dos niveles. En un primer nivel se analizará el DG interalélico positivo entre pares de alelos de microsatélites. Estos DG interalélicos se evaluarán en función de la frecuencia y tamaño de los alelos implicados en cada par, lo que nos permitirá ver si se confirman los patrones de DG dependiente de alelo hallados previamente en la región 11p15 (ZAPATA *et al.* 2001b). En un segundo nivel se analizará el DG global positivo entre pares de microsatélites mediante una medida que resume todos los DG interalélicos positivos en un único valor de significación e intensidad. Se realizará un análisis de la estructura poblacional, lo que será un aspecto importante para la posterior interpretación de los niveles de DG detectados.

Se analizará la distribución del DG, entre pares de SNPs y entre pares de microsatélites, en relación con la frecuencia efectiva de recombinación (FRE) y con la distancia física (DF). Esto nos permitirá investigar si existe

una distancia máxima para la detección del DG en función del marcador genético empleado o si, por el contrario, éste se extiende a lo largo de todo el brazo cromosómico analizado.

La evaluación de los niveles de DG dependiente de marcador nos permitirá determinar qué fracción del desequilibrio detectado es atribuible a las características particulares de cada tipo de marcador genético. Además, será de utilidad para realizar un diseño apropiado de un estudio de localización de genes de enfermedades mediante DG, y escoger el marcador genético más apropiado en cada caso.

- II. El segundo objetivo será investigar la posible implicación de los genes que codifican para el colágeno tipo I (COL1A1 y COL1A2), como factores de riesgo en la otosclerosis. El gen COL1A1 ha sido relacionado con la otosclerosis en base a un estudio de asociación de casos-contrroles (McKENNA *et al.* 1998). En el presente trabajo se pretende realizar un estudio independiente para verificar si se confirman estos resultados previos. Para ello, se evaluará la asociación de la otosclerosis con dos SNPs situados en el gen COL1A1, los cuales habían resultado asociados con la otosclerosis en el estudio de McKENNA *et al.* (1998). Además, se realizará un estudio más exhaustivo del gen COL1A2, mediante un análisis de asociación entre la otosclerosis y seis marcadores polimórficos situados en el gen COL1A2, dos de tipo microsatélite y cuatro de tipo SNP.

Los análisis de asociación se llevarán a cabo mediante un estudio de la intensidad y significación de la asociación comparando las frecuencias alélicas, genotípicas y haplotípicas entre pares de marcadores de una muestra de 100 individuos con otosclerosis y de 100 controles, todos ellos procedentes de la población gallega. Así mismo se llevará a cabo el análisis de asociación estratificando la muestra por sexos. Además, se realizará un análisis de estructura poblacional, lo que será importante para interpretar los resultados y poder descartar falsos positivos en el estudio de asociación.

3. Material y Métodos

I. EVALUACIÓN DEL DESEQUILIBRIO GAMÉTICO DEPENDIENTE DE MARCADOR EN 11p

1. Marcadores genéticos analizados

Para el estudio de los niveles de DG a lo largo del cromosoma 11p se utilizaron dos tipos de marcadores genéticos diferentes, **microsatélites** (repeticiones en tándem de no más de 6 pb) y **SNPs** (*single nucleotide polymorphisms* o polimorfismos de un único nucleótido) con el objeto de comparar los patrones resultantes. Los microsatélites, como se mencionó en el apartado de introducción, son marcadores multialélicos y los SNPs son generalmente bialélicos. En este trabajo se analizaron 18 SNPs bialélicos y 22 loci microsatélites de repetición dinucleotídica (CA)_n distribuidos a lo largo del brazo corto del cromosoma 11 (11p). La figura 2 muestra su localización cromosómica.

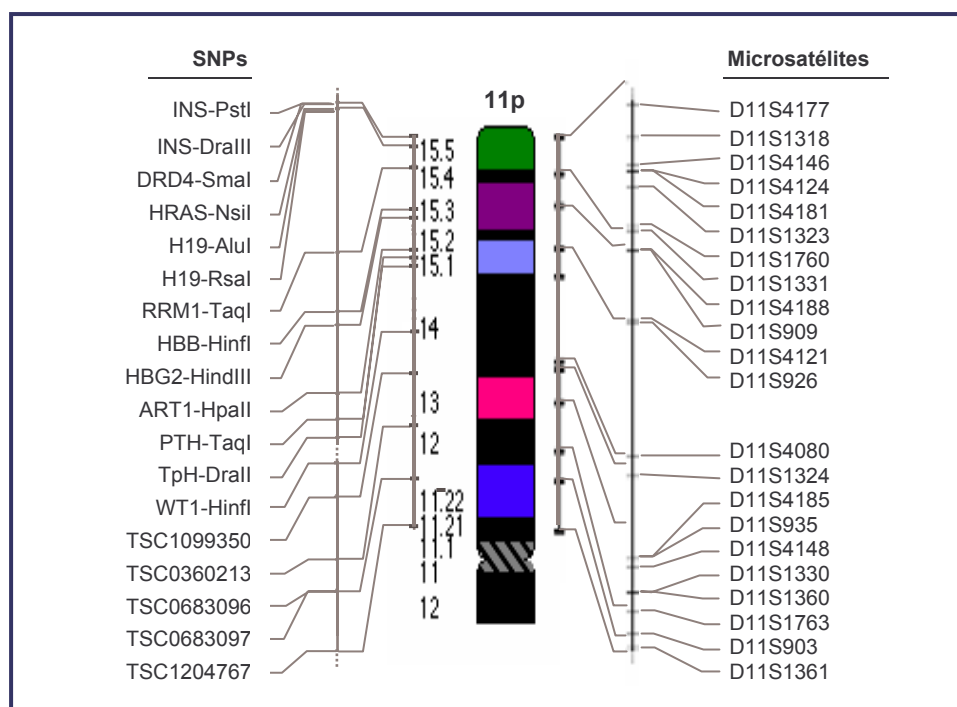


Figura 2. Localización cromosómica de los SNPs y microsatélites estudiados a lo largo del cromosoma 11p.

En la tabla 1 se muestra la relación de los 22 loci **microsatélites** junto con el rango de tamaño y número de repeticiones dinucleotídicas a partir de DIB *et al.* (1996). La nomenclatura de los microsatélites (MCALPINE 1995) tiene el formato de Dn1Sn2, donde D indica que la secuencia está constituida por DNA, n1 indica el cromosoma en el que se haya la secuencia, S indica que se trata de una secuencia simple y n2 indica el número ordinal que tiene asignada la secuencia en el cromosoma.

Tabla 1. Microsatélites analizados en el cromosoma 11p					
Locus	pb	(AC) _n	Locus	pb	(AC) _n
D11S4177	183-213	8-23	D11S926	135-145	20-25
D11S1318	123-145	12-23	D11S4080	134-154	19-29
D11S4146	195-211	14-22	D11S1324	110-128	15-24
D11S4124	165-177	14-20	D11S4185	139-147	14-18
D11S4181	201-217	16-24	D11S935	196-208	12-18
D11S1323	201-207	14-17	D11S4148	174-182	18-22
D11S1760	80-100	10-20	D11S1330	156-160	19-21
D11S1331	191-205	16-23	D11S1360	103-117	15-22
D11S4188	109-123	15-22	D11S1763	182-190	16-20
D11S909	113-125	16-22	D11S903	99-109	18-23
D11S4121	92-122	15-30	D11S1361	206-220	16-23

pb: tamaño en pares de bases, n: número de repeticiones del dinucleótido

Los 18 **SNPs** seleccionados se extienden a lo largo del brazo corto del cromosoma 11, desde la región 11p15 a 11p11. El análisis de estos SNPs se realizó mediante dos metodologías diferentes que describiremos en los siguientes apartados. 13 SNPs fueron analizados por amplificación mediante PCR y posterior tratamiento con enzimas de restricción (PCR-RFLP) y 5 SNPs analizados mediante electroforesis en geles con gradiente de desnaturalización (DGGE).

En la tabla 2 se muestra la relación de SNPs analizados por PCR-RFLP junto con los tamaños de sus alelos y sus respectivas referencias bibliográficas. En el caso de estos SNPs, el polimorfismo bialélico afecta a una diana de restricción de manera que puede ser detectado mediante tratamiento con la

endonucleasa de restricción apropiada. Así uno de los alelos estará representado por una secuencia de DNA con la diana de restricción intacta (alelo +) y el otro alelo aparecerá con esa secuencia alterada en al menos un nucleótido (alelo -), de manera que el enzima no la reconocerá y no será capaz de actuar cortando el fragmento de DNA por el sitio específico. La nomenclatura utilizada en estos SNPs deriva del gen en el que se localiza y de la endonucleasa de restricción que se requiere para la detección del polimorfismo. Debido a la ausencia en la literatura de una regla general para la nomenclatura de SNPs basados en PCR-RFLP se hace difícil su búsqueda en bases de datos y una identificación inequívoca en el caso de estar presentes en diferentes trabajos. Actualmente existen bases de datos fácilmente accesibles de multitud de SNPs con información de frecuencias alélicas, tamaño de muestra analizado, distancias físicas y genéticas y método utilizado para su identificación. Con el objeto de facilitar la localización de los SNPs de este estudio en bases de datos que aporten información adicional para este o para estudios posteriores, se realizó una comparación de las secuencias flanqueantes de cada SNP con secuencias reflejadas en *dbSNP* (NCBI HomePage, <http://www.ncbi.nlm.nih.gov/>). Así, se llegaron a identificar algunos SNPs en esta base de datos y en esos casos se refleja en la tabla 2 también su código que consiste en las letras "rs" más un número asignado según su orden de entrada en la base de datos.

Los SNPs analizados mediante DGGE se seleccionaron de entre los presentes en la base de datos *The SNP Consortium* (TSC, <http://www.snp.cshl.org/>). Su nomenclatura consta de las letras "TSC" y un número según su orden de entrada en la base de datos. En la tabla 3 se muestra la relación de SNPs, junto con sus alelos, analizados mediante DGGE.

Tabla 2. SNPs analizados mediante PCR-RFLP

Locus	Diana de restricción	Alelo (-) (pb)	Alelo (+) (pb)	Referencias
INS-PstI (rs3842752)	CTGCA / G	221	137 + 84	BENNETT <i>et al.</i> 1995
INS-DraIII (rs3842748)	CACNNN / GTG	360	231 + 129	HOBAN y KELSEY 1991
DRD4-SmaI	CCC / GGG	700	300 + 400	PETRONIS <i>et al.</i> 1994
HRAS-NsiI	ATGCA / T	112	87 + 25	HOBAN <i>et al.</i> 1991
H19-AluI	AG / CT	174	130 + 44	YBALLE <i>et al.</i> 1996
H19-RsaI	GT / AC	575	407 + 168	KONDO <i>et al.</i> 1995
RRM1-TaqI	T / CGA	219	164 + 55	BYRNE y SMITH 1993
HBB-Hinfi	G / ANTC	341	244 + 154	SEMENZA <i>et al.</i> 1989
HBG2-HindIII (rs2070972)	A / AGCTT	335	237 + 98	CUI <i>et al.</i> , 1989
ART1-HpaII (rs2280134)	C / CGG	120	61 + 59	YADOLLAHI-FARSANI <i>et al.</i> 1999
PTH-TaqI	T / CGA	196	136 + 60	MULLERSMAN <i>et al.</i> 1992
TpH-DraII	RG / GNCCY	323	182 + 141	PAOLINI-GIACOBINO <i>et al.</i> 2000
WT1-Hinfi (rs1799933)	G / ANTC	171	152 + 19	TADOKORO <i>et al.</i> 1993

R = A o G ; Y = C o T ; N = Cualquier nucleótido

Tabla 3. SNPs analizados mediante DGGE

Locus	Alelos
TSC0360213	C / T
TSC0683096	A / C
TSC0683097	A / G
TSC1099350	C / T
TSC1204767	C / T

1.1. Distancias físicas (DF) y frecuencias de recombinación efectivas (FRE) entre los loci

La DF y la FRE entre los loci son parámetros fundamentales para un estudio de caracterización de los patrones de DG. Es necesario conocer la posición relativa de unos marcadores respecto a otros y relacionar este dato con el desequilibrio existente entre ellos.

1.1.1. DF entre los marcadores genéticos analizados

El cálculo de las distancias físicas entre pares de loci es inmediato a partir de la información obtenida de las bases de datos, ya que estas distancias físicas son aditivas. Así, las distancias físicas entre pares de microsatélites se obtuvieron de COLLINS *et al.* (1996) (LDB, <http://www.cedar.genetics.soton.ac.uk>) y entre pares de SNPs a partir de NCBI (NCBI HomePage, <http://www.ncbi.nlm.nih.gov/>). Las tablas 4 y 5 muestran las distancias físicas entre pares de SNPs y entre pares de microsatélites respectivamente.

Tabla 4. Distancias físicas (Mb) entre pares de SNPs

LOCUS	DRD4-SmaI	HRAS-NsII	H19-AluI	H19-RsaI	INS-PstI	INS-DraIII	ARTI-HpaII	RRM1-TaqI	HBB-HinfI	HBG2-HindIII	PTH-TaqI	TPH-DraII	TSC1099350	WT1-HinfI	TSC0360213	TSC0683096	TSC0683097
HRAS-NsII	0,10																
H19-AluI	1,41	1,31															
H19-RsaI	1,41	1,31	7 10 ⁻⁴														
INS-PstI	1,58	1,47	0,16	0,16													
INS-DraIII	1,58	1,47	0,16	0,16	3 10 ⁻⁴												
ARTI-HpaII	3,08	2,97	1,67	1,66	1,50	1,50											
RRM1-TaqI	3,54	3,43	2,12	2,12	1,96	1,96	0,46										
HBB-HinfI	4,64	4,54	3,23	3,23	3,07	3,07	1,57	1,11									
HBG2-HindIII	4,67	4,57	3,26	3,26	3,09	3,09	3,09	1,13	0,03								
PTH-TaqI	12,99	12,89	11,58	11,58	11,41	11,41	9,91	9,45	8,35	8,32							
TPH-DraII	17,55	17,44	16,13	16,13	15,97	15,97	14,47	14,01	12,90	12,87	4,55						
TSC1099350	27,20	27,10	25,79	25,79	25,63	25,63	24,13	23,67	22,56	22,53	14,21	9,66					
WT1-HinfI	31,85	31,75	30,44	30,44	30,27	30,27	28,77	28,31	27,21	27,18	18,86	14,31	4,65				
TSC0360213	33,22	33,12	31,81	31,81	31,64	31,64	30,14	29,68	28,58	28,55	20,23	15,67	6,02	1,37			
TSC0683096	34,06	33,96	32,65	32,65	32,48	32,48	30,98	30,52	29,42	29,39	21,07	16,52	6,86	2,21	0,84		
TSC0683097	34,06	33,96	32,65	32,65	32,48	32,48	30,98	30,52	29,42	29,39	21,07	16,52	6,86	2,21	0,84	1 10 ⁻⁴	
TSC1204767	36,59	36,48	35,17	35,17	35,01	35,01	33,51	33,05	31,94	31,92	23,60	19,04	9,38	4,74	3,37	2,53	2,52

Tabla 5. Distancias físicas (Mb) entre pares de loci microsatélites

Locus	D11S4177	D11S1318	D11S4146	D11S4124	D11S4181	D11S1323	D11S1760	D11S1331	D11S4188	D11S909	D11S4121	D11S926	D11S4080	D11S1324	D11S4185	D11S935	D11S4148	D11S1330	D11S1360	D11S1763	D11S1361
D11S1318	0,64																				
D11S4146	2,02	1,38																			
D11S4124	2,62	1,98	0,60																		
D11S4181	2,62	1,98	0,60	0,00																	
D11S1323	3,21	2,58	1,19	0,60	0,60																
D11S1760	0,29	0,34	1,72	2,32	2,32	2,92															
D11S1331	8,29	7,65	6,27	5,67	5,67	5,08	7,99														
D11S4188	9,48	8,85	7,47	6,87	6,87	6,27	9,19	1,20													
D11S909	9,52	8,88	7,50	6,90	6,90	6,31	9,23	1,23	0,04												
D11S4121	14,51	13,88	12,49	11,90	11,90	11,30	14,22	6,22	5,03	4,99											
D11S926	14,11	13,47	12,09	11,49	11,49	10,90	13,81	5,82	4,63	4,59	0,40										
D11S4080	28,73	28,69	27,61	26,88	26,88	26,15	24,25	21,07	19,87	19,84	15,24	14,84									
D11S1324	30,35	30,31	29,23	28,50	28,50	27,77	25,87	22,69	21,49	21,46	16,86	16,46	1,62								
D11S4185	37,38	37,35	36,26	35,53	35,53	34,80	32,90	29,72	28,52	28,49	23,89	23,49	8,65	7,03							
D11S935	37,39	37,35	29,23	35,54	35,54	34,81	32,91	29,73	28,53	28,50	23,90	23,50	8,66	7,04	0,01						
D11S4148	37,53	37,49	36,40	35,67	35,67	34,95	33,05	29,86	28,67	28,63	24,04	23,63	8,80	7,18	0,14	0,14					
D11S1330	37,81	37,77	36,69	35,96	35,96	35,23	33,33	30,15	28,95	28,91	24,32	23,92	9,08	7,46	0,43	0,42	0,29				
D11S1360	38,41	38,37	37,29	36,56	36,56	35,83	33,93	30,75	29,55	29,52	24,92	24,52	9,68	8,06	1,03	1,02	0,89	0,60			
D11S1763	41,51	41,48	40,39	39,66	39,66	38,94	37,04	33,85	32,66	32,62	28,03	27,62	12,79	11,17	4,13	4,13	3,99	3,70	3,10		
D11S1361	44,94	44,91	43,82	43,09	43,09	42,37	40,47	37,28	36,09	36,05	31,46	31,05	16,21	14,59	7,56	7,55	7,42	7,13	6,53	3,43	
D11S903	45,50	45,46	44,38	43,65	43,65	42,92	41,02	37,84	36,64	36,61	32,01	31,61	16,77	15,15	8,12	8,11	7,98	7,69	7,09	3,99	0,56

El cromosoma 11 tiene una longitud total de 134 Mb y el brazo corto representa el 40% del total del cromosoma (53,1 Mb). Los marcadores genéticos analizados abarcan una gran parte del brazo corto del cromosoma 11. Así, los microsatélites analizados se extienden a lo largo de 45,50 Mb lo que representa el 85% del brazo corto. Y de la misma manera los SNPs se extienden a lo largo de 36,59 Mb lo que supone el 69% del brazo corto.

En la tabla 6 se muestran los valores promedio y el rango de las DFs entre pares de SNPs y entre pares de microsatélites.

Tabla 6. Descriptivos de las DF (Mb) para pares de SNPs y para pares de loci microsatélites		
	SNPs	Microsatélites
Rango	$1 \times 10^{-4} - 36,59$	0 – 45,50
Media $\pm$ SE	$16,08 \pm 1,05$	$19,82 \pm 0,95$

Los promedios de las DFs son similares en ambos tipos de marcadores, aunque ligeramente inferiores en los SNPs ($16,08 \pm 1,05$ Mb en SNPs vs $19,82 \pm 0,95$ Mb en microsatélites). El menor valor promedio en los SNPs puede ser debido al mayor número de SNPs estudiados estrechamente ligados (3 SNPs separados por $1 \times 10^{-4} - 7 \times 10^{-4}$ Mb) en comparación con los microsatélites (1 microsatélite separado por 0 Mb). Así, el rango varía desde loci estrechamente ligados hasta un valor máximo próximo a la segregación independiente (36,59 Mb en los SNPs y 45,50 Mb en los microsatélites). El valor mínimo de DF que se indica para los microsatélites es una aproximación puesto que estos marcadores son adyacentes en la base de datos, pero en realidad estos microsatélites están separados por unas pocas pares de bases.

1.1.2. FRE entre pares de SNPs y entre pares de microsatélites

Las FRE para los pares de SNPs se calcularon a partir de los datos de distancias genéticas por sexos reflejados en COLLINS *et al.* (1996) (LDB, <http://www.cedar.genetics.soton.ac.uk>) para el caso de los SNPs analizados mediante PCR-RFLP y de MATISE *et al.* (2003) (TSC, http://snp.cshl.org/linkage_maps/) para el caso de los SNPs analizados mediante DGGE. Para la transformación de distancias genéticas en frecuencias de recombinación se utilizó la función de mapa de Kosambi. Este cálculo se hizo por sexos y a continuación se realizó un promedio para obtener la frecuencia de recombinación efectiva. El cálculo de las FRE entre cada par de microsatélites se realizó a partir de los datos de DIB *et al.* (1996) que nos da las frecuencias de recombinación por sexos entre pares de marcadores adyacentes. Estas frecuencias de recombinación no son aditivas de manera que se utilizó la función de mapa de Kosambi para transformarlas en distancias genéticas. Así, a partir de las distancias genéticas se calculó la frecuencia de recombinación entre todos los pares de microsatélites y posteriormente se realizó un promedio entre los datos de los dos sexos.

La función de mapa es una relación matemática entre la distancia genética y la frecuencia de recombinación (STRACHAN & READ 1999). Una de las primeras funciones de mapa que se obtuvieron, fue la de HALDANE (1919)

$$m = -\frac{1}{2} \ln(1 - 2\theta) \quad \rightarrow \quad \theta = \frac{1}{2} (1 - e^{-2m})$$

Donde θ es la fracción de recombinantes y m es la distancia de mapa en centiMorgans. Si los entrecruzamientos ocurriesen aleatoriamente y no influyesen unos en otros esta función de mapa sería perfectamente válida. Sin embargo, esta no es la realidad y la presencia de un quiasma inhibe la formación de otro en las proximidades. Este fenómeno se denomina interferencia de entrecruzamiento o de quiasma y se puede definir como la posición no aleatoria de los entrecruzamientos

en las cromátidas individuales (BROMAN y WEBER 2000). La interferencia (I) fue descrita primeramente por STURTEVANT (1915) y MULLER (1916) poco tiempo después de la aparición de los primeros modelos de recombinación y se calcula como:

$$I = 1 - \text{c.o.c.}$$

Donde c.o.c. es el coeficiente de coincidencia, que es la relación entre los entrecruzamientos observados y los esperados.

Se han desarrollado otras funciones de mapa que tienen en cuenta el fenómeno de interferencia, como por ejemplo ELSESTEIN (1979), OTT (1985), STAHL (1989) y LYON (1980). Pero la de uso más extendido es la función de mapa de Kosambi. KOSAMBI (1914) realizó algunas asunciones específicas sobre la interferencia y derivó una función de uso más extendido que ha dado resultados congruentes con datos observados de eucariotas superiores:

$$m = \frac{1}{4} \ln \frac{(1 + 2\theta)}{(1 - 2\theta)} \quad \rightarrow \quad \theta = \frac{1}{2} \frac{(e^{4m} - 1)}{(e^{4m} + 1)}$$

En las tablas 7 y 8 se muestran las FRE entre cada par de SNPs y entre cada par de microsatélites, respectivamente.

Tabla 7. Frecuencias de recombinación efectivas (cM) entre pares de SNPs

LOCUS	INS-DraIII	INS-PstI	DRD4-SmaI	HRAS-NsiI	H19-AluI	H19-RsaI	RRM1-TaqI	HBB-HinfI	HBG2-HindIII	ARTI-HpaII	PTH-TaqI	TpH-DraII	WT1-HinfI	TSC1099350	TSC0360213	TSC0683096	TSC0683097
INS-PstI	0,00																
DRD4-SmaI	0,01	0,01															
HRAS-NsiI	0,44	0,44	0,43														
H19-AluI	0,45	0,45	0,44	0,01													
H19-RsaI	0,45	0,45	0,44	0,01	0,00												
RRM1-TaqI	9,19	9,19	9,18	8,77	8,76	8,76											
HBB-HinfI	12,77	12,77	12,76	12,36	12,35	12,35	3,80										
HBG2-HindIII	13,54	13,54	13,54	13,14	13,13	13,13	4,61	0,81									
ARTI-HpaII	17,77	17,77	17,77	17,39	17,38	17,38	9,16	5,43	4,63								
PTH-TaqI	19,27	19,27	19,26	18,90	18,89	18,89	10,83	7,14	6,35	1,74							
TpH-DraII	20,36	20,36	20,35	19,99	19,98	19,98	12,05	8,39	7,61	3,03	1,29						
WT1-HinfI	30,94	30,94	30,93	30,67	30,66	30,66	24,58	21,58	20,93	16,93	15,37	14,19					
TSC1099350	31,76	31,76	31,75	31,50	31,49	31,49	25,70	22,82	22,16	18,20	16,69	15,54	2,08				
TSC0360213	34,28	34,28	34,27	34,05	34,04	34,04	28,78	26,13	25,54	21,90	20,48	19,39	5,70	4,34			
TSC0683096	35,16	35,16	35,16	34,95	34,94	34,94	30,00	27,50	26,91	23,34	21,98	20,93	7,76	6,27	1,95		
TSC0683097	35,16	35,16	35,16	34,95	34,94	34,94	30,00	27,50	26,91	23,34	21,98	20,93	7,76	6,27	1,95	0,00	
TSC1204767	37,04	37,04	37,04	36,86	36,85	36,85	32,58	30,42	29,84	26,50	25,28	24,31	12,11	10,75	6,53	4,66	4,66

Tabla 8. Frecuencias de recombinación efectivas (cM) entre pares de loci microsatélites

Locus	D11S4177	D11S1318	D11S4146	D11S4124	D11S4181	D11S1323	D11S1760	D11S1331	D11S4188	D11S909	D11S4121	D11S926	D11S4080	D11S1324	D11S4185	D11S935	D11S4148	D11S1330	D11S1360	D11S1763	D11S5903
D11S1318	2,90																				
D11S4146	5,18	2,30																			
D11S4124	5,88	3,00	0,70																		
D11S4181	5,97	3,10	0,80	0,10																	
D11S1323	7,45	4,59	2,30	1,60	1,50																
D11S1760	10,83	8,04	5,78	5,09	4,99	3,50															
D11S1331	11,50	8,72	6,47	5,78	5,68	4,20	0,70														
D11S4188	13,10	10,36	8,13	7,45	7,35	5,88	2,40	1,70													
D11S909	13,19	10,45	8,23	7,55	7,45	5,98	2,50	1,80	0,10												
D11S4121	19,01	16,47	14,39	13,75	13,65	12,25	8,91	8,23	6,56	6,47											
D11S926	19,09	16,56	14,48	13,84	13,75	12,35	9,00	8,32	6,66	6,56	0,10										
D11S4080	28,03	26,10	23,91	23,34	23,34	22,75	20,40	19,98	18,72	18,68	13,93	13,41									
D11S1324	34,71	28,08	26,03	25,49	25,49	24,93	22,70	22,30	21,09	21,05	16,47	15,97	2,299								
D11S4185	39,33	34,57	32,94	32,54	32,54	32,12	30,28	29,98	29,06	29,03	25,33	24,90	13,00	10,08							
D11S935	39,33	34,57	26,03	32,54	32,54	32,12	30,28	29,98	29,06	29,03	25,33	24,90	13,00	10,08	0,00						
D11S4148	39,46	34,84	33,20	32,81	32,81	32,41	30,57	30,28	29,38	29,35	25,72	25,29	14,02	11,88	1,10	1,10					
D11S1330	40,38	35,86	34,38	33,99	33,99	33,59	31,91	31,62	30,72	30,69	27,16	26,76	17,09	15,03	4,49	4,49	3,40				
D11S1360	40,88	36,90	35,42	35,08	35,08	34,73	33,05	32,80	32,01	31,99	28,72	28,33	17,18	15,12	4,59	4,59	3,50	0,10			
D11S1763	41,97	38,32	36,99	36,68	36,68	36,35	34,84	34,60	33,87	33,85	30,81	30,45	19,43	17,45	7,15	7,15	6,07	2,70	2,60		
D11S903	42,17	38,76	37,44	37,15	37,15	36,84	35,33	35,11	34,42	34,40	31,50	31,14	21,68	19,77	7,15	9,78	8,71	5,38	5,28	2,70	
D11S1361	42,44	39,32	38,01	37,75	37,75	37,47	35,96	35,76	35,15	35,13	32,40	32,04	23,12	21,27	11,50	11,50	10,45	7,15	7,05	4,49	1,70

Como se mencionó previamente para la DF, el brazo corto del cromosoma 11 representa el 40 % del total del cromosoma 11 (58 cM el brazo corto/144 cM todo el cromosoma). En el presente estudio se ha analizado una gran parte del brazo corto del cromosoma 11. Así, los SNPs se extienden a lo largo de 37,04 cM lo que supone el 64% del brazo corto. Y de la misma manera los microsatélites se extienden a lo largo de 42,44 cM lo que representa el 73% del brazo corto.

En la tabla 9 se muestran los valores promedio y el rango de las FRE entre pares de SNPs y entre pares de microsatélites.

Tabla 9. Descriptivos de la FRE (cM) para pares de SNPs y para pares de loci microsatélites		
	SNPs	Microsatélites
Rango	0 – 37,04	0 – 42,44
Media $\pm$ SE	18,25 $\pm$ 0,96	20,06 $\pm$ 0,86

Los promedios de la FRE son similares en ambos tipos de marcadores, siendo ligeramente inferior en SNPs (18,25 $\pm$ 0,96 cM en SNPs vs 20,06 $\pm$ 0,86 cM de FR en microsatélites). El menor valor promedio de FRE en los SNPs (al igual que para la DF) puede ser debido al mayor número de SNPs estudiados estrechamente ligados (3 SNPs frente a 1 microsatélite separados por 0 cM). Así, el rango varía desde loci estrechamente ligados (0 cM en SNPs y en micros) hasta un valor máximo próximo a la segregación independiente (37,04 cM en SNPs y 42,44 cM en micros).

Se observa que las distancias genéticas y las distancias físicas no se corresponden en muchos casos ya que las FR no son uniformes a lo largo de los cromosomas (CHAKRAVARTI *et al.* 1984; GABRIEL *et al.* 2002; TWELLS *et al.* 2003; WALL *et al.* 2003). La aproximación de 1 cM = 1 Mb es una regla útil pero la correspondencia actual varía ampliamente entre diferentes regiones cromosómicas (STRACHAN y READ 1999). En general, las frecuencias de

recombinación son mayores en los telómeros que en los centrómeros (BUETOW *et al.* 1991; WATKINS *et al.* 1994; TIRET *et al.* 2002). La relación promedio entre distancia genética y física en hombres es de 1 cM = 1,05 Mb y en mujeres 1 cM = 0,70 Mb, siendo la media entre los dos sexos de 1,2 Mb/cM, o lo que es lo mismo, 1cM = 0,88 Mb (MORTON *et al.* 1992; KAESSMANN *et al.* 1999; STRACHAN y READ 1999).

2. GENOTIPADO

2.1. Genotipado de los loci microsatélites

Los datos genotípicos para los 22 loci **microsatélites** estudiados provienen de dos trabajos realizados con anterioridad en nuestro laboratorio (HERMIDA 2001; ZAPATA *et al.* 2001). En el trabajo de HERMIDA (2001) se analizaron 10 microsatélites localizados a lo largo de la región centromérica e intersticial del brazo corto del cromosoma 11 (11p14-11) para una muestra de 405 individuos sin parentesco entre sí tomados de la población gallega. Así mismo, en el trabajo de ZAPATA *et al.* (2001) se analizaron 12 microsatélites localizados a lo largo de la región telomérica del brazo corto del cromosoma 11 (11p15) en la misma muestra de 405 individuos. Así, los datos extraídos de estos trabajos se tomaron conjuntamente para el estudio del desequilibrio gamético en 11p.

2.2. Genotipado de los SNPs

2.2.1. Material biológico

Para el estudio del desequilibrio gamético a lo largo del brazo corto del cromosoma 11 humano se obtuvo una muestra de 405 individuos no emparentados entre sí de la población gallega.

De cada individuo se extrajo una muestra de sangre periférica que fue suministrada por el Dr. Carlos Martín perteneciente al Servicio de Otorrinolaringología del Hospital Clínico Universitario de Santiago de Compostela.

Las muestras de sangre se obtuvieron usando EDTA como anticoagulante, transportadas en frío al laboratorio y conservadas a -40° C para su posterior utilización.

2.2.2. Extracción de DNA

La purificación de DNA se ha llevado a cabo utilizando los kits de purificación de *DNA QIAmp®DNA Blood Mini Kit* (Qiagen). El proceso se fundamenta en la realización de una cromatografía de adsorción durante la cual el DNA se queda fijado a la columna. Posteriormente, la adición del tampón de elución, hace que el DNA se libere quedando en suspensión en un tubo Eppendorf con el tampón de elución. Siguiendo el protocolo recomendado por el fabricante, el rendimiento es aproximadamente de 30 ng/μl de DNA partiendo de 200 μl de sangre total. El producto final es una suspensión de DNA que se conservó a -40°.

2.2.3. Genotipado de los SNPs analizados mediante PCR-RFLP

2.2.3.1. Amplificación por la PCR de los SNPs

La amplificación del DNA correspondiente a cada marcador se ha realizado mediante la técnica de la reacción en cadena de la polimerasa (PCR) descrita por MULLIS *et al.* (1986) y SAIKI *et al.* (1988). Las amplificaciones se llevaron a cabo en un termociclador MJ Research (PTC-200 DNA Engine).

En la tabla 10 se exponen las secuencias de los cebadores utilizados para la amplificación de cada SNP.

Tabla 10: Secuencia de los cebadores utilizados para el desarrollo de la PCR

SNPs	Cebadores (5'→3')	Referencias
INS-Dralll	CTGGGGAGAAGTACTGGG GCAGAGCAGGTTCCGGAA	BENNETT <i>et al.</i> 1995
INS-PstI	AAGCGTGGCATTGTGGAACAAT CTGGGAGGGGCTCACAACAGT	HOBAN y KELSEY 1991
DRD4-SmaI	AGATACCAGGTGGACTAGGG ACCTCGGAGTAGACGAAGAG	PETRONIS <i>et al.</i> 1994
HRAS-NsiI	P1 : TCAAGCTTGGCAGGTGGGGCAGGAG P2 : TAGGATCCTGGGCTCGCCCGCAGCA P3 : TGACCATCCAGCTGATCCAGATGCA	HOBAN <i>et al.</i> 1991
H19-AluI	CTTTACAACCACTGCACTACCTG(AC)C GATGGTGTCTTTGACGTTGGGCTGA	YBALLE <i>et al.</i> 1996
H19-RsaI	TACAACCACTGCACTACCTG TGGAATGCTTGAAGGCTGCT	KONDO <i>et al.</i> 1995
RRM1-TaqI	ACTAATGGCAATTCCAA ATTATACCTCCTCAAT	BYRNE y SMITH 1993
HBB-HinfI	TGGATTCTGCCTAATAAAA GGGCCTATGATAGGGTAAT	SEMENZA <i>et al.</i> 1989
HBG2-HindIII	AGTGACTAGTGCTGCAAGAA CTCTGCATCATGGGCAGTGA	CUI <i>et al.</i> , 1989
ART1-HpaII	GAGAGGAAGAGGTGCTGATC AGTTGTAGGTGCTGTGCTTG	YADOLLAHI-FARSANI <i>et al.</i> 1999
PTH-TaqI	GATCTTTCCTGGGAAGAAG GATACCTGCAAAAGACATGG	MULLERSMAN <i>et al.</i> 1992
Tph-Drall	ATGGTACTTACTAGCCTGTG CTGTCTCCACAGTTTTGCC	PAOLINI-GIACOBINO <i>et al.</i> 2000
WT1-HinfI	CACTCGCTCAGCTGTCTTCGG CAAGGACCCAGACGCAGAG	TADOKORO <i>et al.</i> 1993

Durante el transcurso de la experimentación se hicieron algunas modificaciones en el proceso de la PCR para mejorar la amplificación con respecto a las condiciones descritas en la literatura (las referencias para cada SNP se citan en la tabla 10). Se incluyó un 10% de glicerol en la mezcla de reacción para aumentar la eficiencia y especificidad de la amplificación de DNA, lo que ayuda a prevenir la formación de estructuras secundarias durante la reacción de amplificación (LU y NÉGRE 1993). Además, el glicerol a esta concentración es también útil en *buffers* que contienen KCl para revertir los falsos negativos debidos a una amplificación preferencial en la PCR (WEISSENSTEINER y LANCHBURY 1996). También se experimentó con la inclusión de BSA (seroalbúmina bovina) en

la mezcla de reacción con el fin de aumentar el rendimiento de la PCR (NAGAI *et al.* 1998), pero en nuestro caso no se obtuvieron mejoras.

Las reacciones de amplificación fueron similares para todos los loci a excepción de la concentración de los cebadores. Las características particulares de la amplificación del locus HRAS-NsiI se detallan en el siguiente apartado. Todas las reacciones se llevaron a cabo en un volumen final de 25 μ l, conteniendo:

- Tampón comercial (Promega M1906), 50 mM ClK, 10 mM Tris-HCl (pH=9,0 a 25°C) y 0,1% Tritón® X-100.
- Glicerol 10% (v/v)
- 1,5 mM MgCl₂
- La concentración de los cebadores (A y B) se detalla en la tabla 11.
- 0,2 mM de cada deoxinucleótido trifosfato (dATP, dGTP, dCTP y dTTP, Promega U1240)
- 1 unidad de Taq polimerasa (Promega M186E)
- 120 ng de DNA genómico

Tabla 11. Concentraciones de los cebadores para la PCR

Locus	μ M	Locus	μ M
INS-DraIII	1,0	HBB-Hinfl	1,0
INS-PstI	1,0	HBG2-HindIII	1,0
DRD4-SmaI	0,5	ART1-HpaII	2,0
H19-AluI	0,8	PTH-TaqI	1,0
H19-RsaI	1,0	TpH-DraII	0,8
RRM1-TaqI	1,0	WT1-Hinfl	0,5

Las concentraciones de los cebadores A y B para cada locus son las mismas

Los parámetros de los ciclos para la realización de la PCR se encuentran detallados en la tabla 12.

Tabla 12. Parámetros de los ciclos de la PCR para los SNPs analizados por PCR-RFLP

Locus	Parámetros de los ciclos	
INS-DraIII	Paso previo de desnaturalización	95° 3 min
	Desnaturalización	94° 1 min
	“Annealing”	55,2° 1 min
	Elongación	72° 45 seg
	Paso final de elongación	72° 3 min
	Nº de ciclos	30
INS-PstI	Paso previo de desnaturalización	97° 5 min
	Desnaturalización	94° 1 min
	“Annealing”	60° 1 min
	Elongación	73° 2 min
	Paso final de elongación	72° 3 min
	Nº de ciclos	30
DRD4-SmaI	Paso previo de desnaturalización	95° 5 min
	Desnaturalización	95° 1 min
	“Annealing”	56° 30 seg
	Elongación	72° 30 seg
	Paso final de elongación	72° 3 min
	Nº de ciclos	32
H19-AluI	Desnaturalización	95° 40 seg
	“Annealing”	59° 40 seg
	Elongación	72° 40 seg
	Paso final de elongación	72° 4 min
	Nº de ciclos	32
H19-RsaI	Paso previo de desnaturalización	94° 30 seg
	Desnaturalización	94° 1 min
	“Annealing”	64° 3 min
	Elongación	72° 5 min
	Nº de ciclos	30
RRM1-TaqI	Desnaturalización	94° 3 min
	“Annealing”	50° 1,5 min
	Elongación	70° 1 min
	Desnaturalización	92° 45 seg
	“Annealing”	44,9° 1,5 min
	Elongación	70° 1 min
	Nº de ciclos	30
	Desnaturalización	92° 45 seg
	“Annealing”	50° 1,5 min
	Elongación	70° 7 min
HBB-HinfI	Desnaturalización	92° 1 min
	“Annealing”	55° 1 min
	Elongación	72° 2 min
	Paso final de elongación	72° 2 min
	Nº de ciclos	30

Tabla 12. (continuación)

Locus	Parámetros de los ciclos	
HBG2-HindIII	Desnaturalización	95° 30 seg
	"Annealing"	60° 30 seg
	Elongación	72° 1 min
	Paso final de elongación	72° 3 min
	N° de ciclos	30
ART1-HpaII	Desnaturalización	95° 30 seg
	"Annealing"	60° 30 seg
	Elongación	72° 1 min
	Paso final de elongación	72° 3 min
	N° de ciclos	30
PTH-TaqI	Desnaturalización	95° 30 seg
	"Annealing"	60° 30 seg
	Elongación	72° 1 min
	Paso final de elongación	72° 3 min
	N° de ciclos	30
TpH-DraII	Desnaturalización	95° 1 min
	"Annealing"	56° 1 min
	Elongación	72° 1 min
	Paso final de elongación	72° 3 min
	N° de ciclos	30
WT1-HinfI	Desnaturalización	95° 1 min
	"Annealing"	55° 1 min
	Elongación	72° 1 min
	Paso final de elongación	72° 3 min
	N° de ciclos	30

2.2.3.2. Amplificación del locus *HRAS-NsiI*

En el codón 27 del oncogen c-Ha-ras aparece un SNP bialélico previamente descrito como una variante de secuencia (TAPAROWSKY *et al.* 1982). En este caso el genotipado no puede ser directo mediante tratamiento con endonucleasas de restricción, ya que este SNP no está incluido en una diana de restricción. HOBAN *et al.* (1991) hicieron posible su detección diseñando un primer (P3) que es capaz de incorporar la base polimórfica a un sitio de restricción *NsiI*. P3 contiene dos "mismatches" (subrayados en la tabla 10) adyacentes al polimorfismo hacia el final 3' del primer. Así, mediante dos amplificaciones consecutivas empleando los primers P1 y P2 en la primera, y P2 y P3 en la segunda, resulta un fragmento de DNA de 112 pb que posee el sitio polimórfico incluido en la diana de restricción para *NsiI* (ATGCA/T).

1ª PCR → Para un volumen final de 25 µl se añadieron 240 ng de DNA utilizando para ello 0,1 µM de los primers P1 y P2 (las secuencias de los primers se detallan en la tabla 10). El resto de componentes se añadieron a la mezcla de PCR en las mismas concentraciones que para los loci restantes (ver arriba). Los parámetros de los ciclos fueron:

Paso previo de desnaturalización	97° 5 min
Desnaturalización	92° 1 min
“Annealing”	60° 45 seg
Elongación	74° 1 min
Paso final de elongación	72° 2 min
Nº de ciclos	15

2ª PCR → Para un volumen final de 50 µl se añadió 1 µl del producto de la PCR anterior utilizando 0,5 µM de cada primer P2 y P3 (tabla 10). El resto de componentes de la PCR se ajustaron a las mismas concentraciones de la PCR anterior. Los parámetros de los ciclos para este segundo paso, fueron:

Paso previo de desnaturalización	95° 5 min
Desnaturalización	92° 1 min
“Annealing”	55° 1 min
Elongación	74° 1 min
Paso final de elongación	72° 2 min
Nº de ciclos	25

Una vez obtenido el producto de amplificación de 112 pb con el sitio polimórfico incluido en la diana de restricción para *NsiI*, se procede al tratamiento de este producto con la endonucleasa de restricción *NsiI* cuyo procedimiento se verá en el apartado siguiente.

2.2.3.3. Digestión de los productos de amplificación para los SNPs

En un tubo estéril se añadió, para un volumen final de 20 µl:

Agua estéril (<i>Milli-Q</i> , Millipore)	16,3 μ l
Tampón	2 μ l
BSA (seroalbúmina bovina), 10 μ g/ μ l	0,2 μ l
DNA (producto de PCR)	1 μ l
Enzima de restricción, 10u/ μ l	0,5 μ l

La mezcla se incubó durante 5 horas. El tampón y la BSA se suministran por la casa comercial con el enzima de restricción. Los enzimas de restricción utilizados, así como sus tampones correspondientes y la temperatura óptima de incubación son los siguientes:

- *AluI* (Sigma) + Tampón SA 1x (33 mM Tris-acetato, 66 mM Acetato potásico, 10 mM Acetato de magnesio y 0,5 mM DTT; pH 7,9 a 37°C). T^a óptima → 37°C.
- *DraIII* (New England Biolabs) + Tampón 3 1x (100 mM NaCl, 50 mM Tris-HCl, 10 mM MgCl₂ y 1 mM DTT; pH 7,9 a 25°C). T^a óptima → 37°C.
- *Eco0109I* (*DraII*) (Amersham Pharmacia Biotech) + Tampón L 10x (100 mM Tris-HCl, 100 mM MgCl₂ y 10 mM DTT; pH 7,5 a 37°C).
- *HindIII* (Promega) + Tampón E 10x (60 mM Tris-HCl, 1 mM NaCl, 60 mM MgCl₂ y 10 mM DTT; pH 7,5 a 37°C). T^a óptima → 37°C.
- *Hinfl* (Promega) + Tampón B 10x (60 mM Tris-HCl, 500 mM NaCl, 60 mM MgCl₂ y 10 mM DTT; pH 7,5 a 37°C). T^a óptima → 37°C.
- *HpaII* (Promega) + Tampón A 10x (60mM Tris-HCl, 60 mM NaCl, 60 mM MgCl₂ y 10 mM DTT; pH 7,5 a 37°C). T^a óptima → 37°C.
- *NsiI* (Sigma) + Tampón SH 1x (50 mM Tris-HCl, 100 mM NaCl, 10 mM MgCl₂ y 1 mM DTE; pH 7,5 a 37°C). T^a óptima → 37°C.
- *PstI* (Promega) + Tampón H 10x (900 mM Tris-HCl, 500 mM NaCl, 100 mM MgCl₂ y 10 mM DTT; pH 7,5 a 37°C). T^a óptima → 37°C.
- *RsaI* (Promega) + Tampón C 10x (100 mM Tris-HCl, 500 mM NaCl, 100 mM MgCl₂ y 10 mM DTT; pH 7,9 a 37°C). T^a óptima → 37°C.

- *SmaI* (Promega) + Tampón J 10x (100 mM Tris-HCl, pH 7,5, 500 mM KCl, 70 mM MgCl₂ y 10 mM DTT a 37°C). T^a óptima → 25°C.
- *TaqI* (Promega) + Tampón E 10x. T^a óptima → 65°C.

2.2.3.4. Resolución de los genotipos: Electroforesis

Se utilizó la electroforesis horizontal discontinua en geles de poliacrilamida ultrafinos (0,4 mm de grosor) no desnaturalizantes modificada a partir de HAAS *et al.* (1994), para la separación de los fragmentos de restricción en función de su tamaño.

Como control de peso molecular en los geles de electroforesis se usaron dos *ladder*, una de 100 pb (*100 pb DNA ladder*, Promega, G2101) y otra compuesta por los fragmentos producto de la digestión del fago ϕ X174 con *Hinfl* (*DNA ϕ X174/*Hinfl**, Promega, G1761). La primera contiene fragmentos de DNA cuyos tamaños oscilan entre 100 y 1000 pb, más otro fragmento de 1500 pb y la segunda contiene 20 fragmentos de entre 24 y 726 pb.

Estos controles de tamaño son solamente aproximados ya que en condiciones no desnaturalizantes la conformación del DNA provoca que fragmentos del mismo tamaño pero de distinta secuencia puedan migrar en la matriz de poliacrilamida de manera diferente.

2.2.3.4.1. Preparación del gel

La matriz del gel está compuesta de acrilamida y de piperazina que actúa como *crosslinker*. El tamaño de poro viene dado en función de los parámetros T y C que son el porcentaje de acrilamida y piperazina, respectivamente. Estos parámetros definen el gel y vienen dados por las expresiones:

$$T = \frac{A + B}{V} 100$$

$$C = \frac{B}{B + A} 100$$

Donde A y B son los gramos de acrilamida y piperazina respectivamente y V es el volumen de la disolución final.

El tampón del gel consta de:

- ☞ 0,33 M Tris, pH = 9,0
- ☞ 0,11 M Ácido Fórmico
- ☞ 0,024 M Ácido Bórico
- ☞ 6% Glicerol

En estos SNPs se empleó un tamaño de poro de 7,74T6,25C.

La polimerización del gel necesita un catalizador. Por ello se añade PSA (persulfato amónico) a una concentración final de 2,5 M y 1,6 μ l de TEMED (N,N,N',N'-Tetramethilenthilendiamin) por mililitro de solución del gel, que actúan como los catalizadores de la reacción de polimerización. La mezcla se vierte en un soporte de vidrio de 2 mm de grosor tratado con *BindSilane* (Pharmacia Biotech 17-1330-01) al efecto de que la matriz del gel quede fijada al cristal. El gel polimeriza a 37°C en oscuridad durante 20 minutos.

2.2.3.4.2. Sistema electroforético

Una vez polimerizado el gel se realiza la carga de las muestras. Se empapan papeles de fibra de vidrio (*IEF sample appl. Piece*, Pharmacia Biotech 80-1129-46) con la muestra a analizar que se aplican directamente sobre el gel a 0,5 cm del electrodo anódico. En los extremos del gel se disponen tiras de papel secante de 1 cm de ancho (Blotting Paper, Sigma P-7176) empapadas con el tampón del electrodo compuesto por una solución de 1 M TRIS y 0,28 M de ácido bórico. Encima de éstos se apoyan los electrodos de forma que la dirección de corrido es de ánodo a cátodo al estar el DNA cargado negativamente.

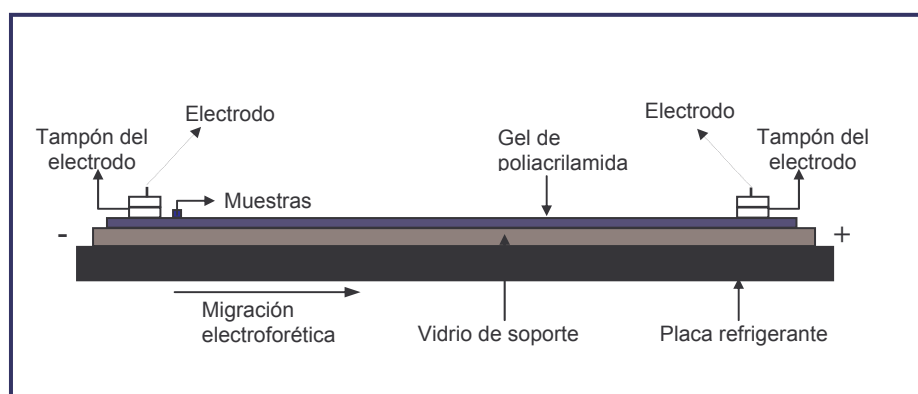


Figura 3. Representación del sistema electroforético empleado para separar los fragmentos de DNA en función de su tamaño.

La migración electroforética se realiza con la placa refrigerante a 15°C en el sistema *Multiphor II* (LKB Pharmacia), representado en la figura 3, con una solución de azul de bromofenol marcando el frente de migración. Las condiciones electroforéticas para todos los loci SNP fueron de 500 V, 35 mA y 20 w; y la migración transcurrió durante 30 minutos con una distancia interelectrodo de 8 cm.

2.2.3.4.3. Tinción del gel

El DNA se visualizó mediante sales de plata (BUDOWLE *et al.* 1991), según el siguiente protocolo.

1. Inmersión del gel en una disolución de etanol al 10% (v/v). El etanol precipita el DNA y de esta forma se adhiere a la matriz del gel.
2. Se sumerge el gel en una disolución de ácido nítrico al 1% (v/v). De esta forma se oxida el DNA y se pone en disposición de unirse al nitrato de plata.
3. Dos lavados con agua *Milli-Q* (Millipore).
4. Tinción con nitrato de plata 0,012 M con suficiente volumen para cubrir el gel completamente.

5. Dos lavados con agua *Milli-Q* (Millipore).
6. Inmersión del gel en una solución de carbonato sódico 0,28 M con 0,06% (v/v) de formaldehído. Esta solución precipita el nitrato de plata de forma que se puede visualizar el DNA.
7. La reacción se para con ácido acético glacial al 10 % (v/v) que fija el gel.
8. Abundantes lavados con agua destilada.

2.2.4. Genotipado de los SNPs analizados mediante DGGE

La técnica DGGE es una herramienta poderosa para detectar pequeñas mutaciones en fragmentos de DNA amplificados por PCR (Wu *et al.* 1999). Esta técnica posee resolución suficiente para distinguir dos moléculas de DNA que se diferencian en un único par de bases (SHEFFIELD *et al.* 1989). Está basada en las propiedades físicas de *melting* de la doble cadena de DNA que dependen de la secuencia de ese DNA y de su estado de desnaturalización. La temperatura de *melting* sería la cual a la que la doble cadena se desnaturaliza. Por ejemplo, dos moléculas idénticas excepto por una posición nucleotídica tendrán propiedades de *melting* similares pero no idénticas. Como resultado de su diferente composición nucleotídica se desnaturalizarán bajo condiciones ligeramente diferentes (diferente concentración de un agente desnaturalizante o diferente temperatura). En la práctica, si estas moléculas se someten a electroforesis con un gradiente positivo de agentes desnaturalizantes (habitualmente urea y formamida), la doble cadena de cada fragmento se separará en posiciones diferentes a lo largo del gel. Además, la tasa de migración de estas moléculas también depende de su estado de desnaturalización de manera que a medida que ésta comienza, su tasa de migración será diferente. El resultado será la diferente posición final de los dos fragmentos en el gradiente desnaturalizante del gel para las mismas condiciones de voltaje y tiempo de corrido electroforético (SHEFFIELD *et al.* 1989; WU *et al.* 1998; HOELZEL 2002; PETERSEN *et al.* 2002).

Para una aplicación exitosa de esta técnica, se deben seleccionar apropiadamente los fragmentos de DNA así como diseñar unos cebadores adecuados (WU *et al.* 1998; HAYES 2003). Para determinar si un fragmento es apropiado para su análisis por DGGE se deben cumplir dos criterios: Primero, se debe prevenir la disociación completa del fragmento lo que se consigue mediante la inclusión de una cola de GCs (guaninas y citosinas). Segundo, la secuencia de DNA debe tener un perfil de *melting* óptimo. Y este óptimo se cumple cuando la secuencia de interés presenta la temperatura de *melting* más baja del total del fragmento amplificado. La introducción de una secuencia rica en GC en el fragmento amplificado no sólo previene la disociación completa del fragmento, sino que incluso altera sus características de *melting* permitiendo la detección de mutaciones en la parte disociada del fragmento. Esta secuencia rica en GCs se incluye a través de la utilización de cebadores en la PCR a los que se les ha añadido una cola de GC en su proceso de síntesis. La inclusión de la cola de GC aumenta el número detectable de variantes de DNA de un 40% a un 100%. Debido a esta elevada fiabilidad, la técnica DGGE se utiliza frecuentemente para la detección de pequeñas mutaciones en enfermedades hereditarias (WU *et al.* 1998).

En el caso que nos ocupa nuestro interés en esta técnica radica en su capacidad para distinguir dos alelos de un SNP en un individuo. Debido al hecho de que un SNP es una variación de una única base, si amplificamos por PCR con los primers apropiados su región flanqueante y lo sometemos a DGGE podremos diferenciar los fragmentos con diferentes alelos por su diferente posición electroforética.

2.2.4.1. Amplificación por PCR de los SNPs analizados mediante DGGE

Los SNPs seleccionados para el estudio se escogieron de entre los descritos en *The SNP Consortium* (TSC, <http://snp.cshl.org/>). Las regiones flanqueantes a los 5 SNPs susceptibles de someterse a DGGE fueron amplificadas mediante PCR en un termociclador MJ Research (PTC-200 DNA

Engine). Las secuencias de los cebadores y sus colas GCs correspondientes se detallan en la tabla 13. Estos cebadores y longitud óptima de la cola de GCs fueron diseñados por Ingeny International (Holanda).

Tabla 13. SNPs analizados mediante DGGE

SNP	Cebadores	GC-Clamp
TSC0360213 SNP2	5'-CTTTTGCCTTCCAGTGTTAC-3' 5'-AGAGTAGTGACTGGCACATA-3'	CGCCCGCCGCGCCCCGCGCCCGGCC CGCCGCCCCCGCCCGCGCCCGCCGC
TSC0683096 SNP3	5'-ATTGGCTGAGTCTAACTGGA-3' 5'-TTCTACCCAGCAAGTCCCTT-3'	CGCCCGCCGCGCCCCGCGCCCGGCC CGCCGCCCCCGCCCGTTTTTTT
TSC0683097 SNP4	5'-TTCACCTGGCCCTCTCTTAC-3' 5'-AATGTTAGTCTCTACTACCC-3'	CGCCCGCCGCGCCCCGCGCCCGGCC CGCCGCCCCCGCCCG
TSC1099350 SNP9	5'-TGGCTGTAGAAAATCTCTCA-3' 5'-TAGGAGAAAAGCCATGAGAA-3'	CGCCCGCCGCGCCCCGCGCCCGGCC CGCCGCCCCCGCCCGTTTTTTT
TSC1204767 SNP10	5'-CTTAACCAAGTCTGACAGTT-3' 5'-TTTAAGGGGAGACGATACAG-3'	CGCCCGCCGCGCCCCGCGCCCGGCC CGCCGCCCCCGCCCGTTTTTTT

Las reacciones de amplificación fueron similares para todos los loci y se llevaron a cabo en un volumen final de 20 µl conteniendo:

- 2 µl de 10x Tampón comercial
- 1,5 mM MgCl₂
- 10 pmol de cada primer
- 0,2 mM de cada deoxinucleótido trifosfato (dATP, dGTP, dCTP, dTTP)
- 1 unidad de Taq DNA polimerasa
- 120 ng de DNA genómico

Los parámetros de los ciclos para la PCR se detallan en la tabla 14.

Tabla 14. Parámetros de los ciclos para PCR previa a DGGE

TSC0360213 TSC0683096 TSC0683097		TSC1099350 TSC1204767	
94° 3 min	5 ciclos	94° 3 min	30 ciclos
94° 1 min		94° 1 min	
56° 1 min		50° 1 min	
72° 2 min		72° 1 min	
94° 1 min	5 ciclos	72° 3 min	
53° 1 min			
72° 2 min			
94° 1 min	25 ciclos		
50° 1 min			
72° 2 min			
72° 5 min			

2.2.4.2. DGGE

Las DGGEs se realizaron en el sistema Ingeny phorU (Ingeny International, Holanda). En la figura 4 se puede ver una fotografía del sistema electroforético.



Figura 4. Fotografía del sistema electroforético para DGGE

2.2.4.2.1. Preparación del gel de electroforesis

Los geles en gradiente de desnaturalización se componen de poliacrilamida al 9%, tampón TAE y dos agentes desnaturalizantes que son urea y formamida. Para la elaboración de los geles de distintos gradientes para el análisis de cada SNP se prepararon dos soluciones stock:

- ❖ Solución stock 1: 80 % urea-formamida (UF), 500 ml de volumen final:
 - TAE 50x: 5 ml
 - Solución del 30 % acrilamida-bisacrilamida (Bio-Rad nº cat. 161-0158): 150 ml
 - Urea: 188 g
 - Formamida: 160 ml
 - Ajustar el volumen hasta 500 ml con agua destilada
- ❖ Solución stock 2: 0 % urea-formamida, 500 ml de volumen final:
 - TAE 50x: 5 ml
 - Solución del 30 % acrilamida-bisacrilamida (Bio-Rad nº cat. 161-0158): 150 ml
 - Ajustar el volumen hasta 500 ml con agua destilada

A partir de estas soluciones se elaboraron los geles para cada SNP. Debajo se exponen los gradientes utilizados en cada caso.

TSC0360213	TSC0683096	TSC0683097	TSC1099350	TSC1204767
10% - 60% UF	0% - 80% UF	0% - 80% UF	20% - 65% UF	20% - 65% UF

Una vez preparado el sistema se procede al llenado del molde del gel. Así, se vierten 24 ml de cada solución en cada reservorio de la pieza formadora del gradiente, yendo la más concentrada en el depósito de salida. La polimerización del gel necesita catalizadores, por ello a estas dos soluciones se les añade justo antes de su utilización 5 µl de TEMED y 50 µl de una solución de APS a una

concentración del 20 %. A continuación se abre el paso de flujo entre los dos depósitos y se enciende la bomba peristáltica que mantiene un paso de flujo constante para el llenado del molde del gel.

Durante la hora de polimerización del gel se prepara el sistema de tampón del electrodo en el que se va a correr la electroforesis de manera que se llena el tanque con 17 l de agua destilada y se le añaden 170 ml de TAE 50x (concentración final del tampón del electrodo TAE 0,5x).

2.2.4.2.2. Carga del gel y migración electroforética

Para la correcta carga de las muestras en los pocillos del gel se requiere la adición de tampón de carga 10x a las muestras. La composición del tampón de carga es: Azul de bromofenol al 0,25% y Glicerol al 30% en agua miliQ. En la tabla 15 se muestran las condiciones electroforéticas para cada locus.

Tabla 15. Condiciones electroforéticas			
Locus	T ^a	Voltaje	Tiempo
TSC0360213	60°	120	18 h
TSC0683096	58°	115	16 h
TSC0683097	58°	115	16 h
TSC1099350	59°	120	16 h
TSC1204767	59°	120	16 h

2.2.4.2.3. Visualización de los resultados

Para la correcta visualización de los patrones de bandas resultantes tras la electroforesis los geles se tiñeron en una solución con bromuro de etidio durante 10 minutos, tras los cuales, se sumergieron en agua destilada durante 20 minutos para eliminar el exceso de bromuro de etidio y así aumentar la nitidez de las bandas.

3. Análisis de los datos

3.1. Análisis de la variabilidad y estructura genética poblacional

La interpretación de los patrones electroforéticos permite conocer la composición genética de cada individuo para cada uno de los loci analizados. Pero a nosotros nos interesa estudiar esos individuos a nivel poblacional, por ello a partir de los datos genotípicos obtenidos se estimaron una serie de parámetros para estudiar la variabilidad y la estructura genética de la población estudiada. El conocimiento de la estructura poblacional, además de tener interés en sí mismo, será fundamental para la estimación de las frecuencias haplotípicas (véase el apartado 3.2.).

3.1.1. Variabilidad genética

Se estimó la heterozigosis observada ($\hat{h}_o$) para cada locus, como el número de heterozigotos dividido por el número total de individuos analizados (N).

La variabilidad genética en un locus se estimó mediante el índice de la heterozigosis esperada ($\hat{h}_e$) o diversidad génica (NEI Y ROYCHOUDHURY 1974; NEI 1978):

$$\hat{h}_e = 1 - \sum_{i=1}^k \hat{p}_i^2$$

Donde p_i es la frecuencia del alelo i estimada a partir de la muestra analizada y k el número de alelos detectados en el locus. Cuando N es pequeño, las estimas de la heterozigosis esperada están influenciadas por el tamaño muestral, por eso es mejor medida de diversidad la estima insesgada de la heterozigosis esperada ($\hat{h}_e$) para cada locus (NEI Y ROYCHOUDHURY 1974; NEI 1978):

$$\hat{h}_e = \frac{2N \left(1 - \sum_{i=1}^k \hat{p}_i^2 \right)}{2N - 1}$$

La estima de la heterozigosis promedio ($\hat{H}$), para un conjunto de r loci analizados, viene dada por la expresión:

$$\hat{H} = \frac{\sum_{i=1}^r \hat{h}_i}{r}$$

Donde $\hat{h}_i$ es la estima insesgada de la heterozigosis esperada para cada locus.

3.1.2. Magnitud y significación de las desviaciones de las proporciones Hardy-Weinberg

3.1.2.1. Magnitud de las desviaciones

En la literatura se han descrito diferentes estimadores para medir la magnitud de las desviaciones genotípicas observadas de las esperadas por Hardy-Weinberg (H-W) (LI Y HORVITZ 1953; CURIE-COHEN 1982; WEIR Y COCKERHAM 1984; CHAKRABORTY Y DANKER-HOPFE 1991). En nuestro caso hemos utilizado el estimador f_T , descrito por ROBERTSON Y HILL (1984), que se basa en la diferencia de homocigotos esperados y observados. El signo de estos coeficientes nos indica la dirección de las desviaciones de manera que un valor de f negativo indica la existencia de un defecto de homocigotos (exceso de heterocigotos) y un signo positivo es indicativo de un exceso de homocigotos. En muestras grandes es el estimador de menor varianza tanto en una población subdividida, como en una población cuyo verdadero valor de f_T sea 0 (ROBERTSON Y HILL 1984).

$\hat{f}_T$ para cada locus se calcula mediante la expresión dada por ROBERTSON Y HILL (1984):

$$\hat{f}_T = \frac{2 \left[\sum_{i=1}^k \frac{\hat{T}_{ii}}{N_i} \right]}{k-1}$$

Siendo N_i el número de alelos i en la muestra, k el número de alelos de cada locus y T_{ii} el estimador insesgado de la desviación de la frecuencia de homocigotos $A_i A_i$ respecto a las proporciones H-W esperadas. Las estimas de $\hat{T}_{ii}$ vienen dadas la expresión:

$$\hat{T}_{ii} = \frac{2(2N-1)N_{ii} - N_i(N_i-1)}{4(N-1)}$$

Siendo N_{ii} el número de individuos homocigotos para el alelo i .

Las estimas de $\hat{f}_T$ varían de -1 a $+1$. Un valor negativo se correspondería con un defecto de homocigotos y un valor positivo con un exceso, siendo $\hat{f}_T = 0$ cuando las proporciones de los genotipos se ajustan a las proporciones H-W.

La varianza de $\hat{f}_T$, para $\hat{f}_T \neq 0$, se calcula mediante la expresión dada por CURIE-COHEN (1982):

$$V(\hat{f}_T) \cong \frac{1-\hat{f}}{2N(k-1)^2} \left\{ 2(k-1) - 2(2k-1)\hat{f} + k^2\hat{f}^2 + \hat{f}(2-\hat{f}) \sum_{i=1}^k \frac{1}{\hat{p}_i} \right\}$$

Cuando el tamaño muestral es grande y $\hat{f}_T = 0$, la varianza se puede aproximar a la expresión (ROBERTSON Y HILL 1984; YASUDA 1968):

$$V(\hat{f}_T) = \frac{1}{N(k-1)}$$

3.1.2.2. Significación de las desviaciones

La significación de las desviaciones de las proporciones H-W para cada locus, se obtuvo aplicando el test χ^2 asociado al estimador $\hat{f}_T$ (ROBERTSON Y HILL 1984), cuya expresión viene dada por:

$$X^2 = \frac{\hat{f}_T^2}{Var(\hat{f}_T)}$$

Para la hipótesis nula $f_T = 0$, este test se distribuye asintóticamente como una distribución χ^2 con un grado de libertad. Tomando como nivel de significación $\alpha = 0,05$, rechazaremos la hipótesis nula cuando X^2 exceda el valor de 3,84.

Se han descrito otros tests para detectar desviaciones en las proporciones H-W, como los tests de probabilidad o tests exactos, el test χ^2 de bondad de ajuste y el test de razón de verosimilitudes (LI Y HORVITZ 1953; YASUDA 1968; EMIGH Y KEMPTHORNE 1975; ELSTON Y FORTHOFFER 1977; EMIGH 1980; LOUIS Y DEMPSTER 1987; HERNÁNDEZ Y WEIR 1989; LESSIOS 1992; RAYMOND Y ROUSSET 1995). Los tests de probabilidad son útiles cuando el tamaño de muestra es pequeño ya que no están basados en la teoría asintótica como los de bondad de ajuste (HERNÁNDEZ Y WEIR 1989; GUO Y THOMPSON 1992). Además, el test exacto es un test más conservador (menor número de errores de tipo I) que el test de bondad de ajuste (LOUIS Y DEMPSTER 1987). Se han propuesto múltiples correcciones de continuidad para hacer los tests de bondad de ajuste válidos para muestras pequeñas (EMIGH Y KEMPTHORNE 1975; ELSTON Y FORTHOFFER 1977), pero

en ocasiones no mejoran mucho los resultados obtenidos con el test de bondad de ajuste convencional (EMIGH 1980; HERNÁNDEZ Y WEIR 1989). Respecto al test de razón de verosimilitudes no se diferencia demasiado en sus resultados del χ^2 de bondad de ajuste convencional. El test utilizado en este trabajo es más potente que el test de bondad de ajuste convencional para el caso multialélico (YASUDA 1968; EMIGH 1980; ROUSSET Y RAYMOND 1995). Además nos permite medir la significación a partir del estimador de la magnitud de las desviaciones que hemos utilizado.

3.2. Estimación del desequilibrio gamético (DG)

El DG designa las asociaciones no al azar entre alelos de loci distintos. Dichas asociaciones pueden estimarse bien partiendo de muestras genotípicas o de muestras gaméticas (HEDRICK 1985). Pero en ambos casos es necesario conocer las frecuencias haplotípicas o gaméticas para la estimación de la intensidad, significación y frecuencia del DG.

3.2.1. Estimación de las frecuencias haplotípicas

Partiendo de una muestra de genotipos no es posible obtener las frecuencias haplotípicas directamente si desconocemos la fase gamética de los individuos dobles heterocigotos. Así, es necesario calcular las estimas de máxima verosimilitud de las frecuencias haplotípicas que maximicen la probabilidad de las frecuencias genotípicas de la muestra. Éste no es un cálculo directo por lo que es necesario utilizar métodos de iteración. Para ello se empleó el algoritmo E-M (Expectación-Maximización) (DEMPSTER *et al.* 1977; EXCOFFIER Y SLATKIN 1995). Hay que tener en cuenta que este algoritmo precisa que las frecuencias genotípicas observadas en la población no se desvíen significativamente de las proporciones H-W ya que cualquier desviación podría generar estimas sesgadas de las frecuencias haplotípicas (EXCOFFIER Y SLATKIN 1995). Existen autores que proponen la utilización del algoritmo E-M incluso cuando algún alelo de los loci del

par no cumple las proporciones H-W, para tamaños de muestra en el rango de moderado a grande (FALLIN & SCHORK 2000).

Los pasos que lleva a cabo el algoritmo EM son los siguientes:

1. Empieza asignando valores arbitrarios a las frecuencias haplotípicas.
2. Estas frecuencias haplotípicas son usadas para inferir la frecuencia de los haplotipos en los individuos dobles heterocigotos, de modo que maximicen la probabilidad de la distribución de los genotipos observados, asumiendo las proporciones H-W.
3. Tras la resolución de las frecuencias haplotípicas en los genotipos dobles heterocigotos se obtienen nuevas estimas de las frecuencias haplotípicas (paso de maximización).
4. Los pasos 2 y 3 se repiten sucesivas veces hasta que los valores converjan entre dos pasos de iteración, que será cuando la diferencia entre la suma de las frecuencias haplotípicas en dos pasos sucesivos no supera el valor de 10^{-7} .

Una vez terminado el proceso, no hay seguridad de que las frecuencias haplotípicas sean las estimas de máxima verosimilitud, ya que se puede dar la existencia de máximos locales. Para evitarlo se debe partir de una gran variedad de condiciones iniciales, sobre todo en muestras pequeñas (DEMPSTER *et al.* 1977; EXCOFFIER Y SLATKIN 1995; LONG *et al.* 1995). En nuestro caso se partió de 100 condiciones iniciales diferentes.

Las estimas de las frecuencias haplotípicas se obtuvieron con el algoritmo EM, con el programa informático *Arlequín ver. 2000* (SCHNEIDER *et al.* 2000).

3.2.2. Estimación de la intensidad del DG

3.2.2.1. Estimación de la intensidad del DG entre loci bialélicos (SNPs)

Consideremos dos loci polimórficos, A y B , con dos alelos cada uno A_1 y A_2 y B_1 y B_2 , respectivamente. Y p_1, p_2, q_1, q_2 las frecuencias alélicas de A_1, A_2, B_1, B_2 respectivamente. Para que el signo de las asociaciones no sea arbitrario, se consideran las clases gaméticas en acoplamiento aquellas formadas por los alelos más o menos frecuentes según LANGLEY *et al.* (1974).

El DG entre pares de loci bialélicos se puede medir mediante el coeficiente D (LEWONTIN & KOJIMA 1960) basado en la diferencia entre la frecuencia observada de un haplotipo y la frecuencia esperada si los alelos se combinaran al azar, que toma la expresión:

$$D = X_1 - p_1q_1$$

siendo X_1 la frecuencia del haplotipo A_1B_1 . Así, p_1q_1 constituye una estima de la frecuencia haplotípica esperada si existiera unión al azar de los alelos de ambos loci.

El valor y el rango de la intensidad del DG estimado mediante coeficientes D es dependiente de las frecuencias alélicas, lo cual hace difícil realizar comparaciones entre pares de loci, dentro de una misma población y entre poblaciones distintas. Para resolver ese problema, LEWONTIN (1964) introdujo un coeficiente de desequilibrio estandarizado cuyo rango es independiente de las frecuencias alélicas y oscila entre -1 y +1 (HEDRICK 1987, 1988; LEWONTIN 1988; ZAPATA & ALVAREZ 1997a).

$$D' = \frac{D}{D_{\max}}$$

donde D_{max} es el máximo valor que D puede tomar, dadas las frecuencias alélicas, y su signo, de manera que:

$$D_{max} = \text{mín} [p_1q_1, p_2q_2] \text{ cuando } D < 0$$

$$D_{max} = \text{mín} [p_1q_2, p_2q_1] \text{ cuando } D > 0$$

Este coeficiente nos permite comparar valores de desequilibrio entre pares de loci que difieren en sus frecuencias alélicas dentro y entre poblaciones (HEDRICK 1987; LEWONTIN 1988; ZAPATA & VISEDO 1995; ZAPATA 2000).

Un inconveniente del coeficiente D' para frecuencias alélicas de valores extremos es que éste alcanza el máximo valor cuando una clase gamética no está presente debido a error de muestreo ($D' = \pm 1$) independientemente del grado de asociación (ZAPATA & ALVAREZ 1992).

La potencia estadística del test chi-cuadrado para detectar DG entre dos loci con dos alelos depende fuertemente del signo de las asociaciones, siendo considerablemente mayor para detectar asociaciones de signo positivo que de signo negativo cuando el desequilibrio no es intenso y las frecuencias alélicas son extremas (BROWN 1975; THOMPSON *et al.* 1988; ZAPATA y ALVAREZ 1993). Además, se ha demostrado que para el caso de dos loci bialélicos la varianza de muestreo de D' es menor para los desequilibrios positivos que para los desequilibrios negativos de la misma intensidad si los haplotipos con los valores más extremos esperados se consideran las clases gaméticas en acoplamiento y las frecuencias alélicas de los dos loci no son 0,5 (ZAPATA *et al.* 1997).

3.2.2.2. Estimación de la intensidad del DG entre pares de loci multialélicos (microsatélites)

El análisis del DG entre pares de microsatélites se realizó a dos niveles diferentes. En un primer nivel se estimó el desequilibrio para cada par de alelos o haplotipos. En un segundo nivel se calculó el DG global, condensando la información del desequilibrio entre todos los alelos del par de loci.

3.2.2.2.1. Estimación de la intensidad de los DG interalélicos

En un sistema multialélico consideramos dos loci polimórficos, A y B, cada uno con alelos A_i ($i = 1, \dots, k$) y B_j ($j = 1, \dots, l$) respectivamente. p_i y q_j son las frecuencias de los alelos i y j respectivamente. El desequilibrio para cada haplotipo de dos loci $A_i B_j$, o desequilibrio interalélico, puede ser considerado separadamente agrupando los datos en A_i vs no A_i (A_{-i}) en el locus A y B_j vs no B_j (B_{-j}) en el locus B. Así, el total de los haplotipos de dos loci se disponen en $k \times l$ tablas de contingencia 2×2 . Las frecuencias de las cuatro clases gaméticas resultantes $A_i B_j$, $A_i B_{-j}$, $A_{-i} B_j$ y $A_{-i} B_{-j}$ serán X_{ij} , $p_i - X_{ij}$, $q_j - X_{ij}$ y $1 - p_i - q_j + X_{ij}$, respectivamente. En este caso las clases haplotípicas en acoplamiento son $A_i B_j$ y $A_{-i} B_{-j}$ (aquellas formadas por los alelos más o menos frecuentes) y las clases haplotípicas en repulsión son $A_{-i} B_j$ y $A_i B_{-j}$.

La extensión del coeficiente D (LEWONTIN & KOJIMA 1960) para un sistema de dos loci multialélicos fue propuesta por WEIR & COCKERHAM (1978) que se define como:

$$D_{ij} = X_{ij} - p_i q_j$$

donde $p_i q_j$ constituye una estima de la frecuencia haplotípica esperada si existiera unión al azar de los alelos de ambos loci.

El valor y el rango de D_{ij} , al igual que ocurría con el coeficiente D es dependiente de las frecuencias alélicas. Así, el coeficiente de disequilibrio estandarizado cuyo rango es independiente de las frecuencias alélicas y oscila entre -1 y +1 (HEDRICK 1987, 1988; LEWONTIN 1988; ZAPATA & ALVAREZ 1997a), toma la expresión:

$$D'_{ij} = \frac{D_{ij}}{D_{ij \max}}$$

donde $D_{ij \max}$ es el máximo valor que D_{ij} puede tomar, dadas las frecuencias alélicas, y su signo, de manera que:

$$D_{ij \max} = \min [p_i q_j, (1 - p_i)(1 - q_j)] \text{ cuando } D_{ij} < 0$$

$$D_{ij \max} = \min [p_i(1 - q_i), q_j(1 - p_j)] \text{ cuando } D_{ij} > 0$$

Este coeficiente nos permite comparar valores de disequilibrio entre pares de loci que difieren en sus frecuencias alélicas dentro y entre poblaciones (HEDRICK 1987; LEWONTIN 1988; ZAPATA & VISEDO 1995; ZAPATA 2000).

El inconveniente del coeficiente D'_{ij} (como ocurría para el D' en el caso bialélico) es que, para frecuencias alélicas de valores extremos, alcanza el máximo valor cuando una clase gamética no está presente debido a error de muestreo ($D'_{ij} = \pm 1$) independientemente del grado de asociación (ZAPATA & ALVAREZ 1992).

Las observaciones acerca de la mayor potencia estadística y menor varianza de muestreo para el DG entre pares de loci con dos alelos se extienden también a los loci multialélicos. Para sistemas multialélicos, si los haplotipos con los alelos más o menos frecuentes son las clases gaméticas en acoplamiento ($A_i B_j$ y $A_{-i} B_{-j}$). Y si $p_i < 0,5$ y $q_j < 0,5$, entonces las asociaciones interalélicas positivas

proporcionan una estimación más precisa de la intensidad del desequilibrio porque las estimas de D'_{ij} con signo positivo tendrán una varianza de muestreo menor que las estimas con signo negativo de la misma intensidad bajo condiciones equivalentes (ZAPATA *et al.* 2001a). Así, las asociaciones negativas tienden a sobreestimar la intensidad del desequilibrio. La mayor potencia estadística para detectar DG positivos depende del hecho de haber definido las clases gaméticas en acoplamiento como aquellas compuestas por los alelos más o menos frecuentes. Si la definición hubiese sido al contrario, la potencia estadística sería mayor en la detección de asociaciones de signo negativo. De esta manera, si consideráramos todo el rango de valores de los desequilibrios interalélicos (negativos y positivos), estaríamos introduciendo un sesgo en nuestros resultados y una infraestima en la proporción de desequilibrio significativo existente en la población de estudio.

Estas observaciones sugieren la conveniencia de estimar la intensidad del desequilibrio considerando solamente los desequilibrios de signo positivo con el fin de obtener unos niveles más precisos de las asociaciones existentes. Por tanto, en la estimación del DG en este trabajo se consideraron sólo los desequilibrios interalélicos positivos ($D'_{ij}(+)$).

3.2.2.2.2. Estimación de la intensidad de los DG globales de signo positivo

La intensidad del desequilibrio global de signo positivo (ZAPATA *et al.* 2001_b) se calcula como:

$$D'(+) = \frac{\sum_{i=1}^k \sum_{j=1}^l p_i q_j D'_{ij}(+)}{\sum_{i=1}^k \sum_{j=1}^l p_i q_j}$$

donde p, q_j son las frecuencias esperadas de los haplotipos con asociación positiva [$D'_{ij}(+)$], asumiendo equilibrio gamético. Esta estima es una medida ponderada. De este modo, los elevados valores de $D'_{ij}(+)$ de los haplotipos escasamente representados en la muestra, contribuirán poco al valor promedio total de $D'(+)$. El rango de $D'(+)$ varía desde 0 a un valor máximo muy cercano o igual a 1 no sólo cuando se cumple que el número de alelos y sus frecuencias son las mismas en los dos loci sino también cuando los alelos más comunes en los dos loci tienen una frecuencia igual o superior a 0,5 o cuando por lo menos un locus del par tiene un alelo con frecuencia superior a 0,5. En los demás casos, aunque no se puede afirmar que el rango sea independiente de las frecuencias alélicas, se ha observado que el máximo valor que puede tomar $D'(+)$ es bastante homogéneo y próximo a 1 (ZAPATA 2000; ZAPATA *et al.* 2001_b).

3.2.3. Significación del DG

3.2.3.1. Significación del DG entre pares de loci bialélicos (SNPs)

La significación de los desequilibrios gaméticos entre pares de SNPs se ha calculado directamente aplicando un test sobre el coeficiente D . La hipótesis nula de $D = 0$ fue testada mediante el estadístico X^2 que toma la expresión:

$$X^2 = \frac{ND^2}{p(1-p)q(1-q)}$$

donde N es el tamaño de la muestra de haplotipos de dos loci (HILL 1974; WEIR 1979; LEWONTIN 1988).

Bajo la hipótesis nula de asociación al azar ($D = 0$), este estadístico, para tamaños de muestra grandes, se distribuye asintóticamente como una χ^2 con un grado de libertad.

3.2.3.2. Significación del DG entre pares de loci multialélicos (microsatélites)

3.2.3.2.1. Significación del DG interalélico positivo

La significación estadística para los DG interalélicos de signo positivo se determinó mediante el estadístico propuesto por WEIR y COCKERHAM (1978) a una cola:

$$X_{ij}^2(+) = \frac{ND_{ij}(+)^2}{p_i(1-p_i)q_j(1-q_j)}$$

siendo N el número de haplotipos de dos loci en la muestra. Véase que esta ecuación es equivalente a la utilizada en la determinación de la significación estadística del DG global en loci bialélicos.

Bajo la hipótesis nula de asociación al azar ($D_{ij}(+) = 0$), este estadístico, para muestras grandes, se distribuye asintóticamente como una χ^2 con un grado de libertad.

Se utilizó el test χ^2 a una cola con la corrección de Yates para evitar un rechazo espúreo de la hipótesis nula de equilibrio gamético cuando el número de gametos esperado es pequeño (EVERITT 1997).

3.2.3.2.2. Significación del DG global positivo

Los desequilibrios gaméticos globales de signo positivo fueron testados para la hipótesis nula de asociación al azar empleando el test chi-cuadrado ajustado para múltiples comparaciones con la corrección de Bonferroni (ZAPATA *et al.* 2001_b). Este método emplea los valores individuales de los tests chi-cuadrado correspondientes a los desequilibrios interalélicos positivos de cada par de loci, para testar la significación del DG global positivo entre ese par de loci. Sin embargo, la existencia de múltiples tests para testar una misma hipótesis requiere

que los diferentes tests sean independientes además de que puede originar un aumento en la tasa de error de tipo I y, por tanto, un exceso de resultados significativos. Para evitar estos problemas se utilizó la corrección de Bonferroni. Esta técnica no requiere que los tests individuales sean independientes, cosa que no ocurre puesto que todas las comparaciones están basadas en el mismo conjunto de datos, permitiendo controlar la probabilidad de rechazar incorrectamente la hipótesis nula (error de tipo I) (RICE 1989).

El nuevo nivel de significación ajustado para múltiples tests, se obtiene dividiendo el nivel de significación predeterminado (α) de 0,05 por el número de tests (N_t). Este nivel de significación corregido ($\alpha' = \alpha/N_t$) determinará el valor a partir del cual un test individual será considerado significativo causando el rechazo de la hipótesis nula para el par de loci considerado. Esta técnica puede inflar el error de tipo II cuando la hipótesis nula es falsa (ROTHMAN 1990; ZERBA *et al.* 1991), por lo que ofrece una estima conservadora de la frecuencia de asociaciones no al azar en la región analizada.

3.3. Relación del DG con la FRE y la DF

Para averiguar si existía una relación entre la FRE y la DF con el DG se realizaron tests χ^2 en tablas de contingencia 2x2, comparando la proporción de asociaciones no aleatorias significativas entre marcadores próximos respecto a los más alejados. Para ello se dividieron los pares de marcadores en dos grupos según la FRE y en dos grupos según la DF: pares de marcadores separados por distancias ≤ 2 cM o 2 Mb (estrechamente ligados) y pares de marcadores separados por distancias > 2 cM o 2 Mb (débilmente ligados). El criterio empleado para separar loci próximos (≤ 2 cM o 2 Mb) de loci alejados (> 2 cM o 2 Mb) ha sido empleado en varias ocasiones para abordar la misma problemática (PETERSON *et al.* 1995; LAAN y PÄÄBO 1997).

En los SNPs la comparación se realizó con la proporción de desequilibrios significativos obtenida, en relación a la FRE y a la DF. En los microsatélites la

comparación se realizó tanto a nivel de los desequilibrios interalélicos positivos, en relación a la FRE y la DF, como a nivel de los desequilibrios globales positivos, en relación a la FRE y la DF.

La comparación entre la cantidad de asociaciones no aleatorias entre loci próximos y loci alejados no nos informa de la tendencia que pueda seguir la intensidad del desequilibrio al aumentar la FRE y la DF. Para averiguar si el DG decae monótonicamente con estos parámetros, la relación de la FRE y la DF con el DG se estudió mediante dos coeficientes de correlación.

- El coeficiente de correlación paramétrico r de Pearson que mide la relación lineal existente entre dos variables
- El coeficiente de correlación no paramétrico de rangos de Kendall (τ) mide la relación entre dos variables independientes, aunque ésta no sea necesariamente lineal. Ordena las variables en rangos de valores y calcula cómo los rangos de una variable se correlacionan con los rangos de la otra variable sin tener en cuenta el valor del incremento.

La probabilidad de la correlación entre la DF y la FRE con el DG se obtuvo mediante el test de Mantel (MANTEL 1967; SOKAL 1979). Este test compara las matrices de dos variables no independientes, la matriz de valores de las FREs y la matriz de intensidades de DG. El test genera una distribución empírica de coeficientes de correlación entre las dos matrices permutando al azar los elementos de una de las dos matrices un número de veces. En el presente estudio fue de 9.999 permutaciones. El nivel de significación se obtiene comparando la correlación obtenida de los datos observados con la distribución empírica.

El test de Mantel se llevó a cabo empleando la versión 2.11T del programa informático NTSYS-pc (ROHLF 1993).

II. ESTUDIO DE ASOCIACIÓN DE LA OTOSCLEROSIS CON LOS GENES DEL COLÁGENO TIPO I, COL1A1 Y COL1A2

1. Material biológico

Para este estudio se seleccionaron 100 individuos afectados por la enfermedad (69 mujeres y 31 hombres) y 100 sanos control, aleatorios (66 mujeres y 34 hombres). Todos ellos sin relaciones de parentesco entre sí y provenientes de la población gallega. Las edades de los individuos afectados por la enfermedad varían entre 14 y 67 años (promedio de 43 años) y los individuos control entre 14 y 79 años (promedio de 44 años). En el estudio de asociación de una enfermedad es importante la edad de aparición de la misma ya que existe el riesgo de incluir a individuos en el grupo control que en realidad van a manifestar esa enfermedad a una edad más avanzada. De esta manera estaríamos introduciendo un sesgo en el estudio que podría falsear los resultados obtenidos. Los criterios diagnósticos para la inclusión en el estudio de los pacientes con otosclerosis, fueron: Presentar una sordera de conducción demostrada por un análisis audiométrico y otosclerosis confirmada quirúrgicamente.

De cada individuo se extrajo una muestra de sangre periférica que fue suministrada por el Dr. Carlos Martín perteneciente al Servicio de Otorrinolaringología del Hospital Clínico Universitario de Santiago de Compostela. Las muestras de sangre se obtuvieron usando EDTA como anticoagulante, transportadas en frío al laboratorio y conservadas a -40° C para su posterior utilización.

1.1. Extracción de DNA

La extracción de DNA a partir de las muestras de sangre de cada individuo se realizó según el protocolo detallado en el apartado 2.2.2. de la sección anterior.

2. Marcadores genéticos analizados

Los marcadores genéticos analizados para el estudio de asociación de la otosclerosis con los genes candidatos (COL1A1 y COL1A2), se muestran en la figura 5. Como se puede observar, los marcadores genéticos se encuentran en dos regiones cromosómicas donde se localizan los genes COL1A1 y COL1A2 que codifican para la procadena $\alpha(I)$ del colágeno tipo I. El gen COL1A1 abarca una distancia de 18 kb en la región cromosómica 17q21.31-q21.32, y el gen COL1A2 se extiende a lo largo de 39 kb de la región cromosómica 7q22.1.

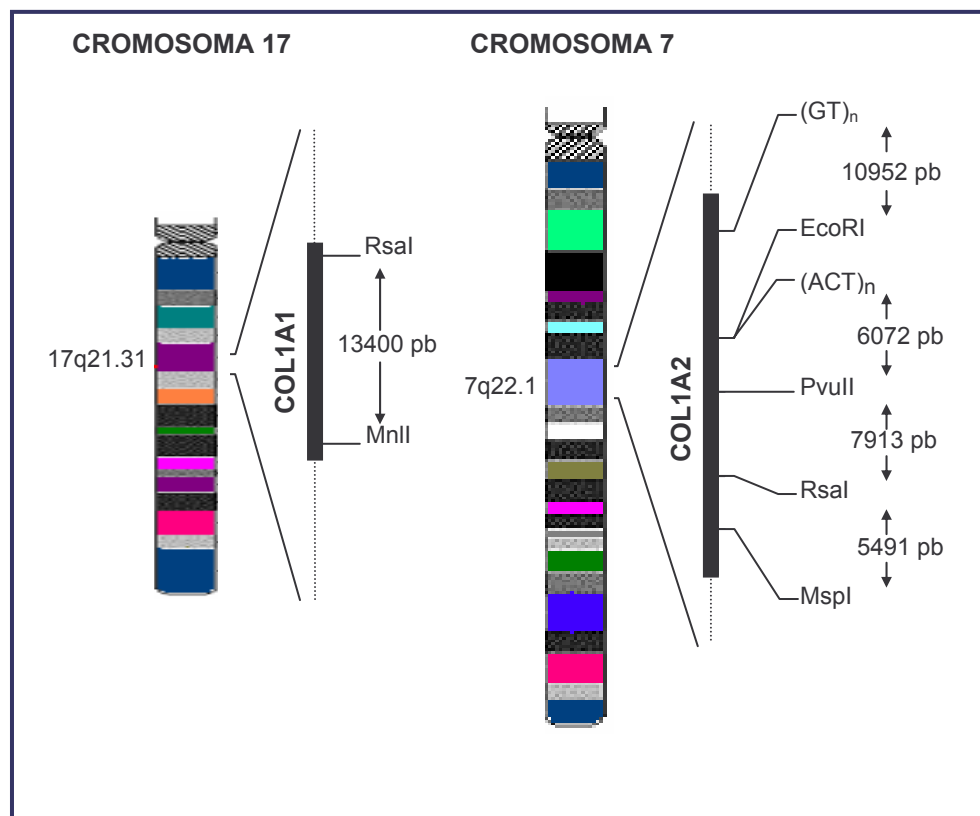


Figura 5. Localización cromosómica de los marcadores genéticos analizados para el estudio de asociación con la otosclerosis. Se muestra la localización citogenética de los genes candidatos, la disposición de los marcadores genéticos dentro de los mismos, así como la distancia física en pb entre cada dos marcadores consecutivos.

La localización de cada marcador dentro del gen, así como el tamaño de sus alelos se muestran en la tabla 16. Los marcadores genéticos analizados fueron dos microsatélites y siete SNPs. Los SNPs son del tipo PCR-RFLP. En estos SNPs el polimorfismo afecta a una diana de restricción de manera que puede ser detectado mediante tratamiento con la endonucleasa de restricción apropiada que reconoce la diana específica y divide la secuencia de DNA en ese punto. Así, uno de los alelos estará representado por una secuencia de DNA con la diana de restricción intacta (alelo +) y el otro alelo aparecerá con esa secuencia alterada en al menos un nucleótido (alelo -), de manera que el enzima no la reconocerá y no será capaz de actuar cortando el fragmento de DNA por el sitio específico. En el SNP MnlI-COL1A1 la nomenclatura de los alelos (T/C) deriva de la base nucleotídica que posee el carácter polimórfico en la diana de restricción. La presencia del alelo T provoca la alteración de la secuencia de la diana de restricción y el alelo C mantiene intacta su secuencia. Las referencias bibliográficas de cada marcador se encuentran especificadas en la tabla 17.

Tabla 16. Marcadores genéticos analizados dentro de los genes COL1A1 y COL1A2 para el estudio de asociación con la otosclerosis

Marcador	Locus (localización)	Tipo de polimorfismo	Tamaño de los alelos (pb)
RsaI-COL1A1	COL1A1 (extremo 5')	SNP	211 + 113 (+) 324 (-)
MnlI-COL1A1	COL1A1 (3' UTR)	SNP	195 (T) 134 (C)
(GT)-COL1A2	COL1A2 (Intrón 1)	Microsatélite	250 252 254 256 258 260
EcoRI-COL1A2	COL1A2 (Intrón 12)	SNP	~170 + 100 (+) ~270 (-)
(ACT)-COL1A2	COL1A2 (Intrón 12)	Microsatélite	6 ^a 7 ^a 8 ^a 9 ^a 10 ^a

Tabla 16. (continuación)

Marcador	Locus (localización)	Tipo de polimorfismo	Tamaño de los alelos (pb)
PvuII-COL1A2	COL1A2 (Codón 392)	SNP	550 + 490 (+) 1040 (-)
RsaI-COL1A2	COL1A2 (Intrón 38)	SNP	800 + 100 (+) 900 (-)
MspI-COL1A2	COL1A2 (Intrón 47)	SNP	850 + 600 + 450 (+) 1300 + 600 (-)

^a Número de repeticiones del trinucleótido

3. Amplificación por la PCR de los marcadores genéticos

La amplificación de los marcadores genéticos se realizó mediante la técnica de la reacción en cadena de la polimerasa (PCR) descrita en el apartado 2.2.3.1. de la sección anterior. Las secuencias de los cebadores utilizados para las amplificaciones se muestran en la tabla 17.

Tabla 17. Secuencia de los cebadores utilizados para la amplificación por PCR de los marcadores utilizados para el estudio de asociación con la otosclerosis

Locus	Cebadores (5'→3')	Referencias
RsaI-COL1A1	GAACTCTCACATACCCTTCACAGT ACTAGAGACATGATGATTAAATGCA	SOKOLOV <i>et al.</i> 1991
MnII-COL1A1	AGACCAGGAATTCGGCTTCG TTGGATCCAAGGTTGAATGCACTTTTGG	SOKOLOV <i>et al.</i> 1991
(GT)-COL1A2	GTTGTTCCACCCACACAGC CTTCATTTTATTCTCTAGGAGCCG	CHI <i>et al.</i> 1992
EcoRI-COL1A2^a (ACT)-COL1A2^a	GGATCCAAAGTCACACATCTAGAG CAATCTATATTCTTATCCTG	PEPE <i>et al.</i> 1993 PEPE <i>et al.</i> 1995
RsaI-COL1A2	CTGCTGGAAGTCGTGGTGAT CACCAGGGAACCAAGTCATA	PEPE <i>et al.</i> 1995
PvuII-COL1A2	GGGATCCTCGGCCCCGCTGAAAAGAA CCGAATTCACCTTTATCACCGTTTTTGCCA	CONSTANTINOI <i>et al.</i> 1990
MspI-COL1A2	TGTAAGTTTGTGAATACCAGT CATCAACTTCATAGTCCTTGG	BAKER <i>et al.</i> 1991

^a Los cebadores son los mismos para la amplificación de estos dos loci

Las mezclas de reacción fueron básicamente las mismas a las descritas en el apartado 2.2.3.1. de la sección anterior, a excepción de la concentración de los cebadores que aparece detallada en la tabla 18.

Tabla 18. Concentraciones de los cebadores utilizados para la amplificación por PCR de los marcadores utilizados para el estudio de asociación con la otosclerosis

Locus	Cebador	μM	Locus	Cebador	μM
RsaI-COL1A1	A	0,8	PvuII-COL1A2	A	0,1
	B	0,7		B	0,1
MnII-COL1A1	A	1,6	RsaI-COL1A2	A	0,14
	B	1,5		B	0,14
(GT)-COL1A2	A	1,3	MspI-COL1A2	A	0,5
	B	1,5		B	0,5
(ACT)-COL1A2	A	1,1			
EcoRI-COL1A2	B	1,5			

La concentración de cebadores que se utilizó para la amplificación mediante la PCR de los SNPs de mayor peso molecular (RsaI-COL1A2, PvuII-COL1A2 y MspI-COL1A2) fue sensiblemente menor que las condiciones estándar debido a que en estos casos no se lograba señal de amplificación trabajando con las condiciones estándar. Trabajos anteriores habían observado una mayor eficiencia y especificidad en la amplificación de fragmentos grandes de DNA cuanto menor concentración de cebadores estuviera presente en la mezcla de reacción (SÁNCHEZ *et al.* 1998; RICHIE *et al.* 1999).

Las reacciones de amplificación se realizaron en un termociclador Perkin Elmer (GeneAmp PCR system 2400). Los parámetros de los ciclos de amplificación mediante la PCR para cada locus, se muestran en la tabla 19.

Tabla 19. Parámetros de los ciclos de amplificación mediante la PCR para los marcadores analizados para el estudio de asociación con la otosclerosis

Locus	Parámetros de los ciclos	
RsaI-COL1A1	Desnaturalización	94°C 1 min
	"Annealing"	55°C 1 min
	Elongación	72°C 1 min
	Rampa de temperatura	100%
	Nº de ciclos	35
MnII-COL1A1	Desnaturalización	94°C 1 min
	"Annealing"	51°C 1 min
	Elongación	72°C 1 min
	Paso final de elongación	72°C 3 min
	Rampa de temperatura	100%
	Nº de ciclos	35
(GT)-COL1A2	Desnaturalización	94°C 1 min
	"Annealing"	55°C 1 min
	Elongación	72°C 2 min
	Paso final de elongación	72°C 2 min
	Rampa de temperatura	100%
	Nº de ciclos	30
EcoRI-COL1A2 y (ACT)-COL1A2	Desnaturalización	95°C 1 min
	"Annealing"	*51°C 2 min
	Elongación	72°C 1 min
	Paso final de elongación	72°C 2 min
	*Rampa de temperatura	40%
	Nº de ciclos	35
PvuII-COL1A2	Desnaturalización	94°C 1,5 min
	"Annealing"	52°C 1 min
	Elongación	73°C 1,5 min
	Rampa de temperatura	100%
	Nº de ciclos	5
	Desnaturalización	94°C 1,5 min
	"Annealing"	64°C 1 min
	Elongación	73°C 1,5 min
	Rampa de temperatura	100%
	Nº de ciclos	30
RsaI-COL1A2	Desnaturalización	95°C 1 min
	"Annealing"	58°C 2 min
	Elongación	72°C 1 min
	Rampa de temperatura	100%
	Nº de ciclos	30
MspI-COL1A2	Paso previo de desnaturalización	94°C 3 min
	Desnaturalización	93°C 0,5 min
	"Annealing"	55°C 1,5 min
	Elongación	70°C 1 min
	Paso final de elongación	72°C 2 min
	Rampa de temperatura	100%
	Nº de ciclos	30

4. Digestión de los productos de amplificación para los SNPs

La digestión de los productos de amplificación para los SNPs se realizó en las mismas condiciones que las detalladas en el apartado 2.2.3.3. de la sección anterior.

Los enzimas de restricción utilizados, así como sus tampones correspondientes y la temperatura óptima de incubación son los siguientes:

- *EcoRI* (Promega) + Tampón H 10x (900 mM Tris-HCl, 500 mM NaCl, 100 mM MgCl₂ y 10 mM DTT; pH 7,5 a 37°C). T^a óptima → 37°C.
- *MnII* (New England Biolabs) + Tampón 2 1x (50 mM NaCl, 10 mM Tris-HCl, 10 mM MgCl₂ y 1 mM DTT; pH = 7,9 a 25°C). T^a optima → 37°C.
- *MspI* (Promega) + Tampón B 10x. T^a óptima → 37°C.
- *PvuII* (Promega) + Tampón B 10x (60 mM Tris-HCl, 500 mM NaCl, 60 mM MgCl₂ y 10 mM DTT; pH 7,5 a 37°C). T^a óptima → 37°C.
- *RsaI* (Promega) + Tampón C 10x (100 mM Tris-HCl, 500 mM NaCl, 100 mM MgCl₂ y 10 mM DTT; pH 7,9 a 37°C). T^a óptima → 37°C.

5. Resolución de los genotipos: Electroforesis

Se utilizó la electroforesis horizontal discontinua en geles de poliacrilamida ultrafinos (0,4 mm de grosor) no desnaturalizantes modificada a partir de HAAS *et al.* (1994) para la separación de los fragmentos de DNA en función de su tamaño (RODRIGUEZ *et al.* 2001; RODRIGUEZ y ZAPATA 2002).

En la tabla 20 se encuentran detalladas las condiciones electroforéticas aplicadas para cada marcador analizado así como el tamaño de poro del gel de poliacrilamida preparado para cada caso (%T %C). Para la puesta en marcha de la electroforesis y su tinción posterior véase el apartado 2.2.3.4. de la sección anterior.

Tabla 20. Tamaño de poro, distancia interelectrodo y condiciones electroforéticas para cada marcador analizado para el estudio de asociación con la otosclerosis

Locus	%T %C ^a	Distancia interelectrodo	Voltaje	Amperaje	Tiempo
(GT)-COL1A2	6,67T 7,14C	22 cm	800 V	25 mA	4 h 30 min
(ACT)-COL1A2	6,67T 7,14C	22 cm	800 V	25 mA	4 h 30 min
EcoRI-COL1A2	7,74T 6,25C	8 cm	500 V	35 mA	30 min
PvuII-COL1A2	7,74T 6,25C	8 cm	500 V	35 mA	30 min
RsaI-COL1A2	6,67T 7,14C	18 cm	500 V	25 mA	2 h
MspI-COL1A2	6,67T 7,14C	18 cm	800 V	25 mA	2 h
MnII-COL1A1	7,74T 6,25C	8 cm	500 V	35 mA	30 min
RsaI-COL1A1	7,74T 6,25C	8 cm	500 V	35 mA	30 min

En todos los casos la potencia se fijó en 20 w.

^a T y C representan el porcentaje de acrilamida y piperazina , respectivamente

Los SNPs se sometieron a electroforesis posteriormente a su tratamiento con el enzima de restricción apropiado. En cambio, en el caso de los microsatélites los fragmentos resultantes del proceso de PCR se someten directamente a electroforesis. En este sistema, los polimorfismos de tipo microsatélite muestran, además de las bandas alélicas, una serie de bandas específicas de alelo y bandas específicas de genotipo que permiten reducir los errores de genotipado de un heterocigoto como homocigoto debido a una no amplificación total o parcial producida durante la PCR (MANIATIS *et al.* 1975; PRINZ *et al.* 1996; RODRÍGUEZ *et al.* 2001).

Las bandas específicas de alelo se dividen en bandas tartamudas o *stutter bands* y bandas diagnóstico o *diagnostic shadow bands* (DSBs). Las bandas tartamudas tienen una mayor tasa de migración que el alelo y aparecen a intervalos de dos pares de bases desde la banda alélica. Las DSBs tienen una menor tasa de migración que el alelo correspondiente y su distancia con la banda alélica es constante dentro de cada locus. Se ha demostrado que las DSBs y los alelos correspondientes son dos conformaciones diferentes de la misma secuencia

(RODRÍGUEZ *et al.* 2001). Las bandas específicas de genotipo se dividen en *heteroduplex bands* y *faint bands*. Las primeras ocurren en individuos heterocigotos y las segundas en homocigotos y poseen una tasa de migración menor que el resto de bandas anteriores.

Además de reducir errores de genotipado al considerar en la lectura las bandas adicionales, este sistema permite una sola base de resolución (RODRÍGUEZ & ZAPATA 2002). Estas bandas asociadas son más frecuentes en los microsatélites de tipo dinucleotídico que en los trinucleotídicos, lo cual se ha podido verificar durante la experimentación.

6. Análisis de los datos

6.1. Análisis de la variabilidad y estructura genética poblacional

El conocimiento de la estructura poblacional es importante en estudios de asociación para investigar si las muestras de individuos controles y enfermos proceden de una población homogénea. Ya que el fenómeno de subestructuración de poblaciones puede originar asociaciones espúreas. Estos análisis se realizaron en las muestras de enfermos, controles y en la muestra total, siendo la metodología estadística empleada similar a la expuesta en el apartado 3.1. de la sección anterior. La única salvedad es el mayor número de alelos que presentan los loci microsatélites respecto a los SNPs que son bialélicos. Por esta razón, las desviaciones de las proporciones H-W para cada alelo se estimaron mediante el estadístico de ROBERTSON Y HILL (1984) para cada alelo ($\hat{f}_{ii}$), que se calcula como:

$$\hat{f}_{ii} = \frac{4NT_{ii}}{N_i(2N - N_i)}$$

Así mismo, su significación se obtuvo aplicando el test χ^2 asociado al estimador $\hat{f}_{ii}$ (ROBERTSON Y HILL 1984). La razón entre la estima $\hat{f}_{ii}$ al cuadrado y su varianza, da lugar al estadístico X_{ii}^2 :

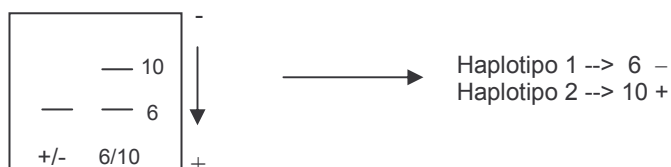
$$X_{ii}^2 = \frac{\hat{f}_{ii}^2}{Var(\hat{f}_{ii})}$$

Para la hipótesis nula $f_{ii} = 0$, este estadístico se distribuye asintóticamente como una χ^2 con un grado de libertad. Tomando como nivel de significación $\alpha = 0,05$, rechazaremos la hipótesis nula cuando X^2 exceda el valor de 3,84.

6.2. Estimación de las frecuencias haplotípicas

Las frecuencias haplotípicas para cada par de loci se estimaron mediante el algoritmo EM (Expectación-Maximización) tanto en la muestra de enfermos como en el grupo control, tal como se describe en el apartado 3.2.1. de la sección anterior.

Para el par de loci EcoRI-COL1A2 y (ACT)-COL1A2 se pudo conocer la fase gamética en los dobles heterocigotos dado que se localizan en el mismo fragmento amplificado. Tras el tratamiento con el enzima de restricción, el alelo (-) del SNP mantiene el tamaño del fragmento amplificado original. Dado que el alelo (-) del SNP posee la misma secuencia que el alelo del microsatélite en su misma fase gamética, ambos comparten la misma tasa de migración electroforética. Los dos fragmentos de menor tamaño resultantes de la digestión con el enzima de restricción y que conforman el alelo (+) del SNP tendrán una mayor migración electroforética que el fragmento amplificado original y por tanto migrarán más que el alelo del microsatélite en su misma fase gamética. En el siguiente esquema se representa un hipotético patrón resultado de una electroforesis al correr los dos marcadores en paralelo.



6.3. Análisis de asociación

El análisis de asociación se realizó mediante la comparación de dos grupos de individuos, una muestra de pacientes afectados de otosclerosis y una muestra de controles.

Consideremos dos loci, en un locus un alelo que predispone para la enfermedad y un alelo normal y en el segundo locus, dos o más alelos de un marcador segregando para ese locus. El genotipo del locus marcador lo averiguamos mediante los métodos experimentales anteriormente descritos. Pero el locus de la enfermedad no podemos visualizarlo ya que no lo conocemos, sólo tenemos datos del fenotipo enfermo/no enfermo. Este fenotipo lo vamos a considerar como un hipotético locus con dos alelos enfermo/no enfermo. De esta manera ya tenemos dos loci entre los que calcular la intensidad de la asociación y su significación. De esta manera los datos se reducen a una tabla 2 x 2 para los marcadores de dos alelos o m x 2, donde m es el número de alelos en un marcador multialélico.

Se realizó un estudio de la intensidad y significación de la asociación comparando las distribuciones de las frecuencias alélicas, genotípicas y haplotípicas entre pares de marcadores de la muestra de enfermos y de la muestra control. Igualmente se llevó a cabo el estudio de asociación estratificando la muestra por sexos, debido a la diferente incidencia de la enfermedad observada en ambos sexos (GORDON 1989; BEN ARAB *et al.* 1993).

6.3.1. Estimación de la intensidad de la asociación y su significación

La intensidad de la asociación entre la otosclerosis y los marcadores localizados dentro de los genes COL1A1 y COL1A2, se estudió mediante el coeficiente D' por ser una medida útil para la localización de genes por desequilibrio (DEVLIN & RISCH 1995; DEVLIN *et al.* 1996; ZAPATA & VISEDO 1995). Existen en la literatura otros coeficientes para medir la asociación entre los marcadores genéticos y las enfermedades en estudios de casos-controles, que varían en sus propiedades y sensibilidad (HEDRICK 1987; DEVLIN & RISCH 1995). DEVLIN & RISCH (1995) realizaron una comparación del comportamiento de los coeficientes o índices de asociación respecto a la frecuencia de recombinación y concluyeron que δ , la versión robusta del riesgo poblacional atribuible, y D' son las mejores medidas para análisis de asociación en estudios de casos-controles. Así mismo, JORDE *et al.* (1994) presentaron resultados de una relación negativa significativa entre el coeficiente D' y la distancia física en la región del gen APC. Existe una ventaja adicional, que no presentan los demás coeficientes, que hace al coeficiente D' apropiado para este estudio y es el buen conocimiento que se posee de sus propiedades para sistemas multialélicos (ZAPATA 2000; ZAPATA *et al.* 2001). El análisis con el coeficiente D' permitirá realizar comparaciones de asociación entre estudios dado que su rango es independiente (para loci bialélicos) o casi independiente (para loci multialélicos) de las frecuencias alélicas (HEDRICK 1987, 1988; LEWONTIN 1988; ZAPATA & ALVAREZ 1997a).

Para el cálculo del coeficiente D' se ha seguido el mismo procedimiento que para la estimación del desequilibrio gamético en 11p. El test de significación de la hipótesis nula de asociación aleatoria entre la otosclerosis y los marcadores bialélicos y los multialélicos se realizó mediante estadísticos chi-cuadrado en tablas de contingencia 2×2 y $m \times 2$, respectivamente, donde m es el número total de alelos en los marcadores multialélicos.

4. Resultados

I. EVALUACIÓN DEL DESEQUILIBRIO GAMÉTICO DEPENDIENTE DE MARCADOR EN 11P

1. ESTUDIO DEL DG ENTRE PARES DE SNPs

1.1. INTERPRETACIÓN DE LOS PATRONES ELECTROFORÉTICOS DE LOS SNPS

1.1.1. SNPs analizados mediante PCR-RFLP

Las técnicas experimentales utilizadas para el genotipado de los individuos requieren someter los productos amplificados por PCR a digestión con endonucleasas de restricción y posterior separación de los fragmentos de DNA resultantes mediante electroforesis. Recordemos que en estos marcadores los alelos se distinguen por la presencia (alelo +) o ausencia (alelo -) de la diana de restricción para una endonucleasa particular.

En la figura 6 se muestran los patrones electroforéticos resultantes de los distintos genotipos para los 13 SNPs analizados por PCR-RFLP. En la parte inferior de la figura se indica el genotipo de cada individuo y en el lateral derecho los tamaños en pb de cada una de las bandas electroforéticas. En los SNPs INS-PstI, INS-DraIII, DRD4-SmaI, H19-RsaI, HBG2-HindIII, PTH-TaqI y TpH-DraII se observan tres fragmentos de DNA de distinto tamaño. La banda electroforética correspondiente al fragmento de DNA de mayor tamaño es el alelo (-). Por su parte, el alelo (+) se evidencia por la presencia de las dos bandas de menor tamaño. Estos dos fragmentos de menor tamaño se generan al estar presente la diana de restricción, de manera que el enzima la reconoce y actúa dividiendo el fragmento amplificado original en dos fragmentos menores. En estos SNPs la diana de restricción ocupa una posición lateral de manera que las dos bandas que se generan son de distinto tamaño y aparecen diferenciadas en el gel de electroforesis. También los SNPs HRAS-NsiI, H19-AluI, RRM1-TaqI y WT1-Hinfi presentan la diana de restricción en una posición lateral pero el fragmento de

Resultados

menor tamaño en el alelo (+) no aparece en la electroforesis al ser demasiado pequeño (25 pb, 44 pb, 55 pb y 19 pb, respectivamente). En cambio, en el caso del SNP ART1-HpaII los fragmentos resultantes de la digestión del alelo (+) con el enzima de restricción son de tamaño similar y se resuelven como una sola banda en la electroforesis.

El caso del SNP HBB-Hinfl es algo diferente a los anteriores. Aquí aparecen 5 bandas en los individuos heterocigotos (+/-), 3 en los homocigotos (-/-) y 4 en los homocigotos (+/+) (figura 6). De la amplificación de la región que contiene este SNP resulta un fragmento de 739 pb que contiene tres dianas de restricción para el enzima Hinfl. Dos de las dianas son constantes y aparecen en todos los individuos de modo que dividen el fragmento de 739 pb en tres de 341 pb, 244 pb y 154 pb. El fragmento de 341 pb es el que contiene la diana de restricción polimórfica de modo que cuando está presente (alelo +), el enzima de restricción lo divide en dos fragmentos de 213 pb y 128 pb. En el alelo (-) se evidencia la banda de 341 pb sin dividir.

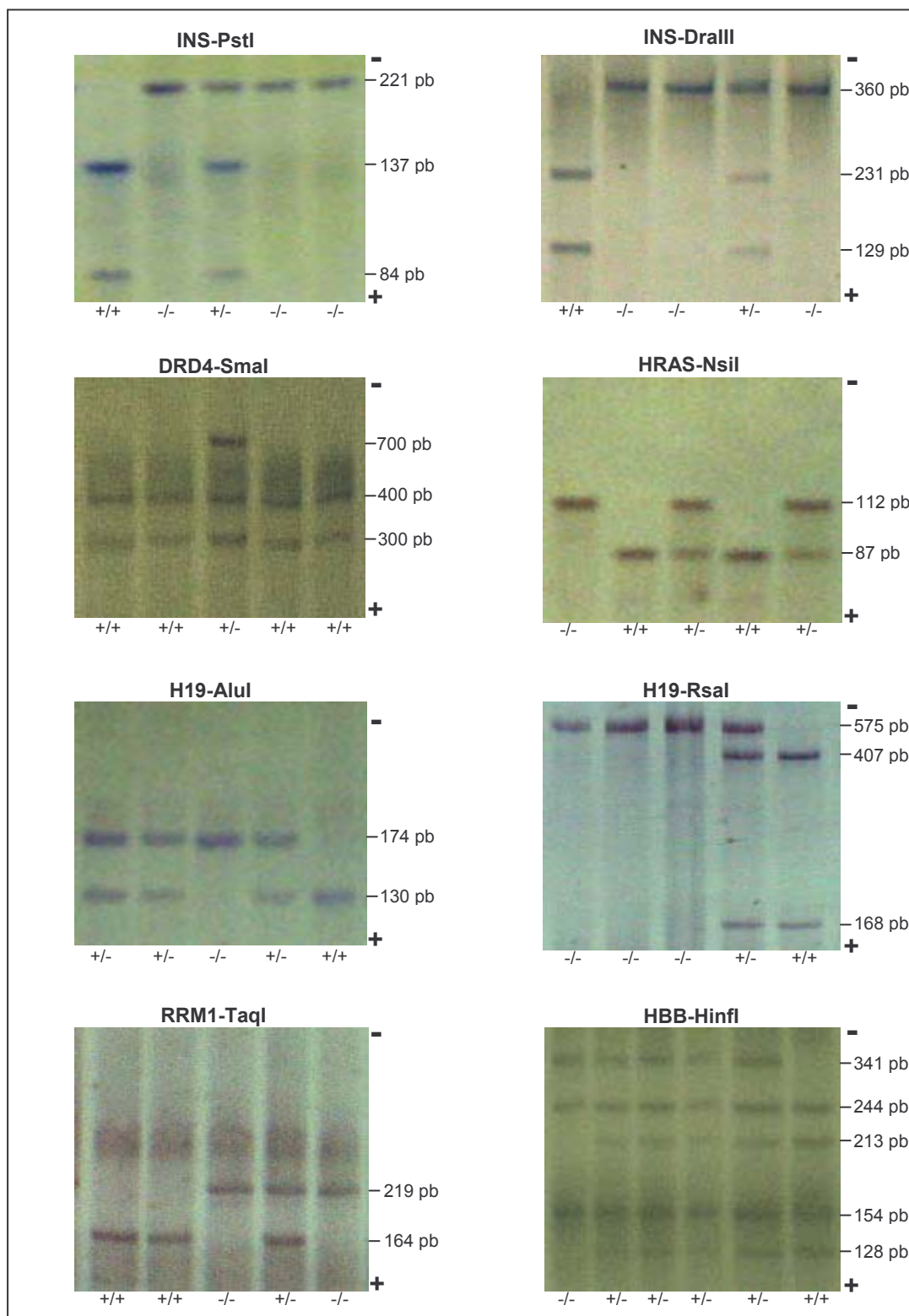


Figura 6. Patrones electroforéticos para los 13 SNPs analizados por PCR-RFLP

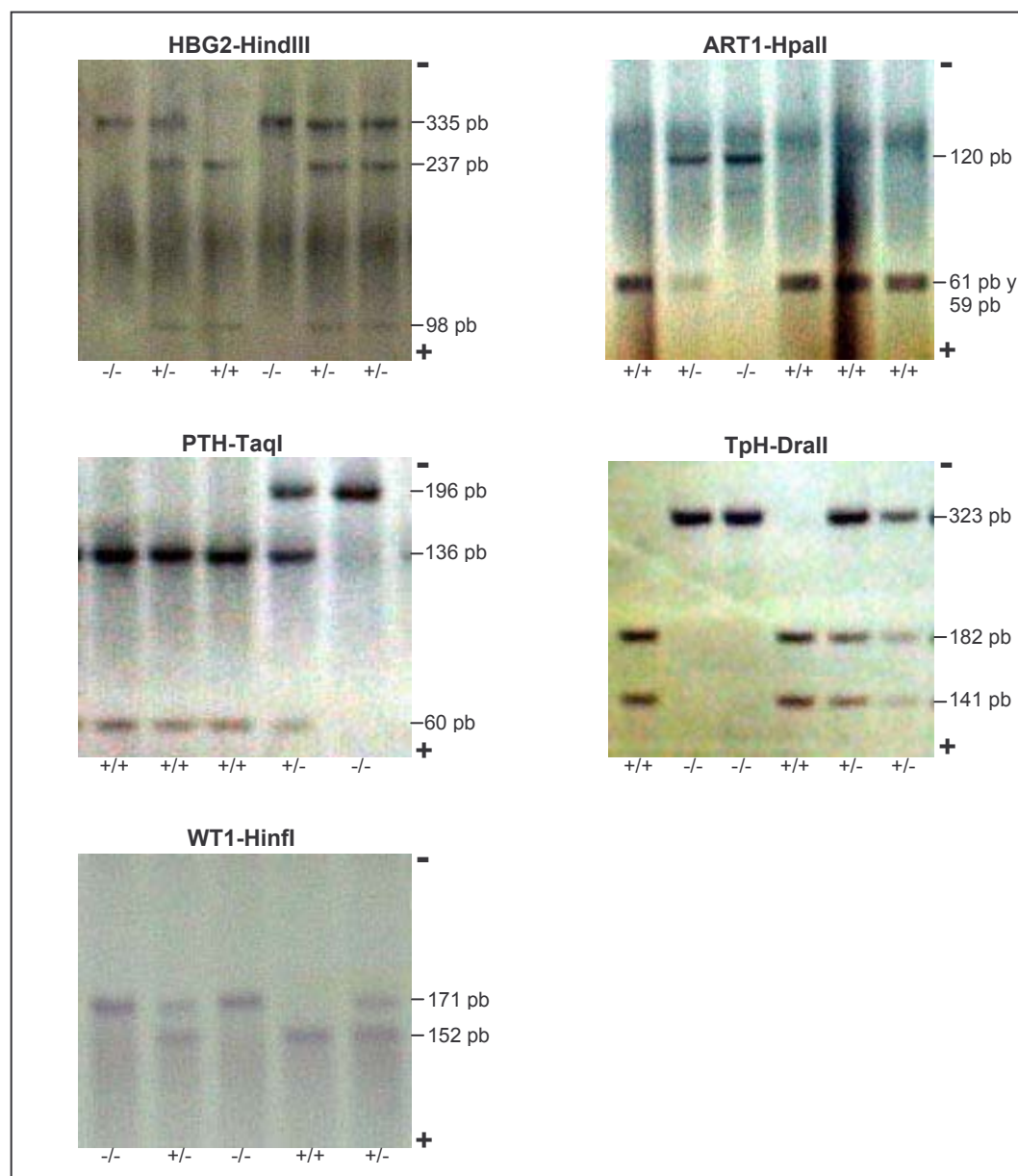


Figura 6. (continuación)

1.1.2. SNPs analizados mediante DGGE

La técnica DGGE nos ha permitido diferenciar dos moléculas de DNA que difieren en su secuencia en una única posición nucleotídica. Los dos alelos de cada SNP se resolvieron en un gel de electroforesis con gradiente de desnaturalización dando lugar a bandas de DNA con diferente tasa de migración electroforética. En la figura 7 se pueden ver los patrones electroforéticos resultantes de los 5 SNPs analizados por DGGE. Cada figura se corresponde con una porción de un gel para uno o dos SNPs.

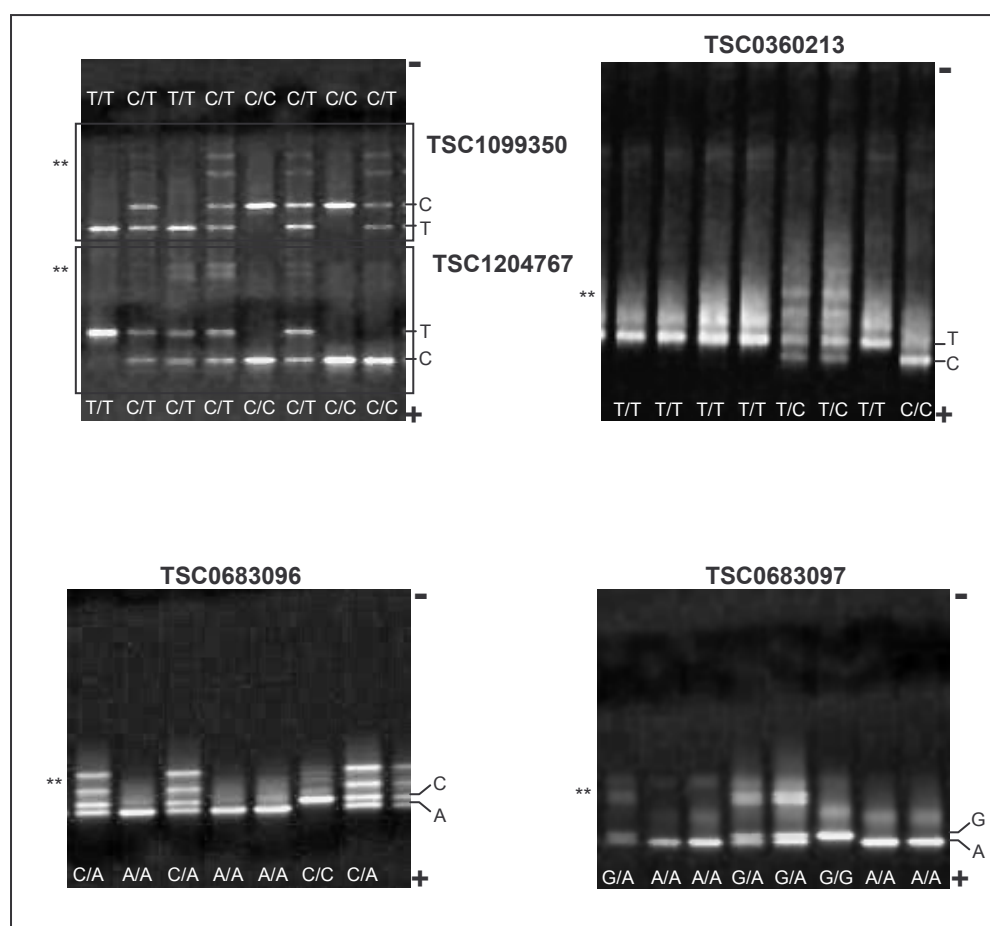


Figura 7. Patrones electroforéticos de los SNPs analizados mediante DGGE. En el lateral derecho se representan los alelos y en la parte inferior los genotipos de cada individuo.

Resultados

Los SNPs TSC1099350 y TSC1204767 se resolvieron en el mismo gel de electroforesis. El área de lectura de cada uno de estos dos SNPs está marcada con un recuadro. Para todos los SNPs en los individuos heterocigotos se observan, además de las bandas alélicas (homodúplex), dos bandas con menor tasa de migración señaladas con dos asteriscos **. Éstas son las bandas heterodúplex que se generan en los últimos pasos de la PCR al estar presentes los dos alelos en heterocigosis.

1.2. ANÁLISIS DE LA VARIABILIDAD Y LA ESTRUCTURA POBLACIONAL

A partir de los datos genotípicos obtenidos se calcularon las frecuencias alélicas para cada SNP. En la figura 8 se muestra la distribución de las frecuencias alélicas observadas en la muestra para cada SNP. En la tabla 21 se muestran las frecuencias absolutas y relativas de cada alelo. Se observa que en casi todos los casos las frecuencias alélicas toman valores intermedios. Sólo en dos SNPs la frecuencia relativa del alelo minoritario fue menor de 0,20 (DRD4-SmaI y WT1-HinfI).

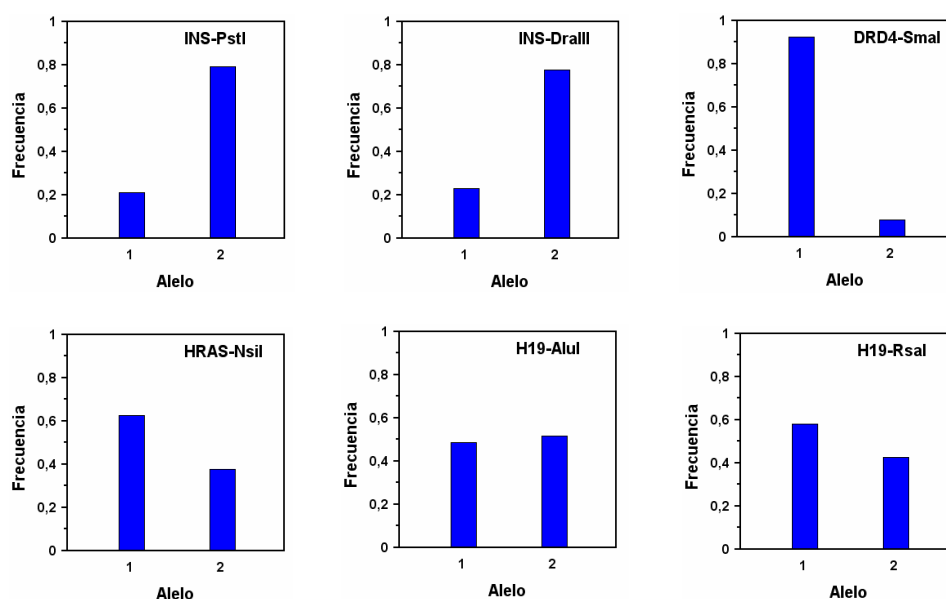


Figura 8. Frecuencias alélicas de los SNPs analizados

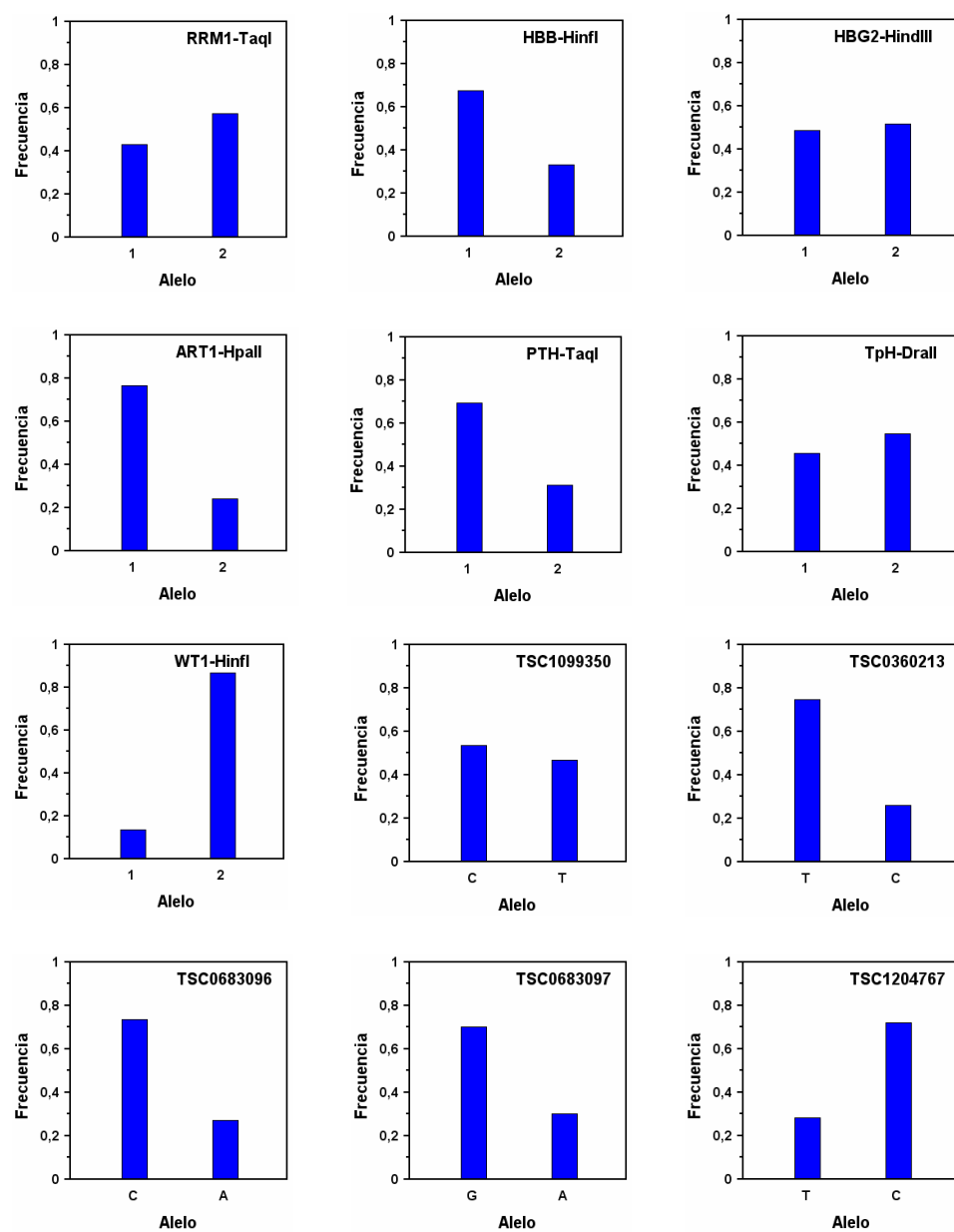


Figura 8. (continuación)

Tabla 21. Frecuencias alélicas de los loci SNPs

Locus	Alelo	Frec.abs.	Frec.rel.
INS-PstI	1 (+)	170	0,210
	2 (-)	640	0,790
INS-DraIII	1 (+)	183	0,226
	2 (-)	627	0,774
DRD4-SmaI	1 (+)	749	0,925
	2 (-)	61	0,075
HRAS-NsiI	1 (+)	506	0,625
	2 (-)	304	0,375
H19-AluI	1 (+)	396	0,489
	2 (-)	414	0,511
H19-RsaI	1 (+)	469	0,579
	2 (-)	341	0,421
RRM1-TaqI	1 (+)	348	0,430
	2 (-)	462	0,570
HBB-Hinfi	1 (+)	544	0,672
	2 (-)	266	0,328
HBG2-HindIII	1 (+)	392	0,484
	2 (-)	418	0,516
ART1-HpaII	1 (+)	617	0,762
	2 (-)	193	0,238
PTH-TaqI	1 (+)	559	0,690
	2 (-)	251	0,310
TpH-DraII	1 (+)	369	0,456
	2 (-)	441	0,544
WT1-Hinfi	1 (+)	108	0,133
	2 (-)	702	0,867
TSC1099350	C	432	0,533
	T	378	0,467
TSC0360213	T	603	0,744
	C	207	0,256
TSC0683096	C	593	0,732
	A	217	0,268
TSC0683097	G	567	0,700
	A	243	0,300
TSC1204767	T	228	0,281
	C	582	0,719

En la tabla 22 se muestran las estimas de variabilidad para cada locus analizado, así como la magnitud y la significación de las desviaciones de las proporciones H-W para cada locus obtenidas mediante el estimador f_T y el test chi-cuadrado, respectivamente (ROBERTSON y HILL 1984). La heterozigosis observada oscila entre 0,15 (DRD4-Smal) y 0,52 (HBG2-HindIII) con un promedio ($\pm$ SE, *standard error*) de $0,40 \pm 0,02$. La heterozigosis esperada (h_e) oscila entre 0,14 (DRD4-Smal) y 0,50 (HBG2-HindIII) que es el máximo valor posible que se puede dar para un sistema de dos alelos. La diversidad, medida en términos del promedio de las heterozigosis esperadas para cada locus, fue de $0,41 \pm 0,02$, lo que sugiere una elevada diversidad para los SNPs analizados en la población de estudio.

Las heterozigosis observada y esperada muestran en general una buena concordancia con un promedio de heterozigosis observada de $0,40 \pm 0,02$ y un promedio de heterozigosis esperada de $0,41 \pm 0,02$. Así mismo en la mayoría de los casos no existen diferencias estadísticamente significativas entre las frecuencias genotípicas observadas y las esperadas bajo las proporciones Hardy-Weinberg. Sólo en el SNP TSC1204767 se detecta una desviación significativa de las proporciones H-W, con una probabilidad ligeramente por debajo del nivel de significación nominal ($P = 0,049$). Este resultado podría ser debido a un error de tipo I (probabilidad de rechazar la hipótesis nula de apareamiento aleatorio cuando es verdadera). Así, el nivel de significación nominal de 0,05 supone que podríamos esperar un 5 % de desviaciones significativas solamente por azar cuando los tests son independientes. Y este porcentaje es aproximadamente el que resulta en este estudio ya que de los 18 tests realizados sólo uno resulta significativo. En este caso los tests no son independientes, por lo que incluso podríamos esperar un mayor porcentaje de casos significativos solamente por azar. Además, en general, la magnitud de las desviaciones de las proporciones H-W, medidas mediante estadísticos f , son pequeñas. Así, los valores oscilan (en valor absoluto) entre $0,001 \pm 0,05$ (Tph-Drall) y $0,098 \pm 0,051$ (TSC1204767), con un promedio de $0,039 \pm 0,006$. Tampoco se observa una tendencia hacia un exceso de homocigotos (f de signo positivo) o defecto de homocigotos (f de signo

Resultados

negativo), ya que el número de casos positivos es equivalente al número de casos negativos.

Tabla 22: Magnitud ($\hat{f}_T$) y significación (χ^2) del ajuste a las proporciones Hardy-Weinberg para cada uno de los loci analizados. P representa la probabilidad obtenida mediante el test Chi-cuadrado.

Locus	h_o	h_e	$\hat{f}_T \pm SE$	$\chi^2(g.l.=1)$	P
INSDraIII	0,368	0,350	-0,051 $\pm$ 0,047	1,040	0,308
INS-PstI	0,346	0,332	-0,041 $\pm$ 0,047	0,684	0,408
DRD4-SmaI	0,146	0,139	-0,045 $\pm$ 0,035	0,813	0,367
HRAS-NsiI	0,440	0,469	0,064 $\pm$ 0,050	1,659	0,198
H19-AluI	0,499	0,500	0,003 $\pm$ 0,050	0,004	0,948
H19-RsaI	0,462	0,488	0,054 $\pm$ 0,050	1,190	0,275
RRM1-TaqI	0,474	0,491	0,034 $\pm$ 0,050	0,467	0,494
HBB-HinfI	0,425	0,442	0,039 $\pm$ 0,050	0,600	0,439
HBG2-HindIII	0,523	0,500	-0,047 $\pm$ 0,050	0,888	0,346
ART1-HpaII	0,388	0,363	-0,067 $\pm$ 0,047	1,806	0,179
PTH-TaqI	0,432	0,428	-0,009 $\pm$ 0,050	0,033	0,856
Tph-DraII	0,496	0,497	0,001 $\pm$ 0,050	0,000	0,988
WT1-HinfI	0,232	0,231	-0,003 $\pm$ 0,049	0,004	0,951
TSC1099350	0,509	0,498	-0,021 $\pm$ 0,050	0,172	0,678
TSC0360213	0,353	0,381	0,073 $\pm$ 0,051	2,180	0,140
TSC0683096	0,378	0,393	0,038 $\pm$ 0,051	0,591	0,442
TSC0683097	0,427	0,421	-0,016 $\pm$ 0,049	0,102	0,750
TSC1204767	0,365	0,405	0,098 $\pm$ 0,051	3,885	0,049
MEDIA $\pm$ S.E.	0,404 $\pm$ 0,023	0,407 $\pm$ 0,023	0,039 $\pm$ 0,006		

1.3. ANÁLISIS DEL DESEQUILIBRIO GAMÉTICO

El análisis del desequilibrio gamético se realizó para todas las combinaciones de pares de loci entre los 18 SNPs analizados del brazo corto del cromosoma 11, lo que supone un total de 153 pares. En este estudio, los SNPs

analizados poseen en su mayoría frecuencias alélicas intermedias por lo cual se optó por considerar todos los desequilibrios independientemente de su signo (positivo o negativo), lo cual no supone una merma en la potencia estadística para la detección del DG. Como se mencionó en el apartado de *Material y Métodos*, existe una mayor potencia estadística para detectar desequilibrios gaméticos de signo positivo que de signo negativo sólo cuando las frecuencias alélicas de los loci implicados se desvían significativamente de las intermedias (BROWN 1980; ZAPATA *et al.* 1997; ZAPATA *et al.* 2001b).

1.3.1. DG entre pares de SNPs

En la figura 9 se muestra la distribución de las frecuencias de la intensidad del DG observado entre pares de SNPs, medido en términos del coeficiente D' . En la tabla 23 se muestran los estadísticos descriptivos y el ajuste a la distribución normal de los valores de D' .

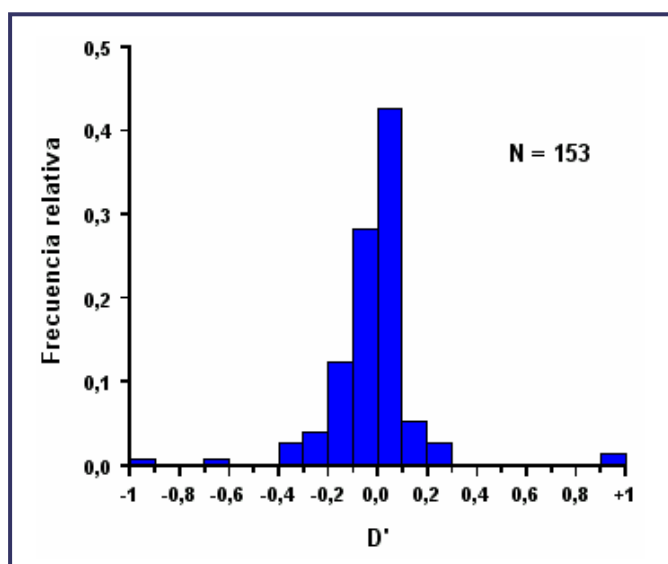


Figura 9. Distribución de frecuencias de los D' observados, donde N es el número de pares de SNPs analizados

Tabla 23. Estadísticos descriptivos, valores de sesgo y curtosis y ajuste a la distribución normal ($D_{(KS)}$) de las distribuciones de los D' . N indica el número de pares de loci.

	SNPs	
	D'	$ D' $
N	153	153
Rango	-0,982 – 0,985	0,0002 – 0,985
Media $\pm$ SE	-0,015 $\pm$ 0,015	0,106 $\pm$ 0,012
Varianza	0,034	0,023
Sesgo (g_1)	0,71 $\pm$ 0,20	
Curtosis (g_2)	15,18 $\pm$ 0,39	
$D_{(KS)}$	0,183***	

***P < 0,001

Se observa que los valores de D' abarcan casi todo el rango posible de intensidades, con un mínimo de -0,982, un máximo de 0,985 y una media ($\pm$ SE) de -0,015 $\pm$ 0,015. Si tomamos estos datos en valor absoluto se obtiene un promedio de intensidad débil, $|D'| = 0,106 \pm 0,012$, con un mínimo próximo a 0 y un máximo de 0,985. El test de Kolmogorov-Smirnov muestra que la distribución de D' observada no se ajusta a la distribución normal ($P < 0,001$). La distribución muestra un sesgo positivo lo que nos indica una falta de asimetría en la distribución, con una gran acumulación de valores de D' en el centro de la distribución (distribución leptocúrtica).

En la tabla 24 se muestran los pares de loci con DG significativo, junto con el valor de la probabilidad obtenida mediante el test χ^2 con y sin la corrección de Yates. También se muestra el valor de D' para los pares de loci. Se observa que sólo un 6,5 % (10/153) de las asociaciones entre los pares de loci son estadísticamente significativas, tanto con como sin la corrección de Yates. La concordancia en el número de casos significativos se explica por el elevado polimorfismo y tamaño de muestra de la población de estudio. La excepción es el SNP DRD4Smal con una frecuencia del alelo minoritario menor de 0,1. Sin

embargo, el par DRD4Smal x HRASNsil es significativo tanto con como sin la corrección de Yates, lo que apoyaría que no estamos ante un falso positivo de asociación no al azar. Además, si aplicáramos la corrección de Bonferroni para determinar la significación estadística de las asociaciones, todavía tendríamos tres casos significativos, con y sin la corrección de Yates, (INSPstI x INSDraIII, H19Alul x H19Rsal y TSC0683096 x TSC0683097). En este estudio no es necesario aplicar un test tan conservador puesto que, como se verá más adelante, disponemos de otras evidencias que sugieren que los desequilibrios detectados no son espúreos. Se observa que no todos los SNPs participan en asociaciones significativas. Así, de los 18 SNPs analizados sólo 11 están incluidos en algún par significativo y, entre ellos, el SNP H19Alul está implicado en el 40 % (4/10) de todos los pares de SNPs significativos. Entre los casos significativos el par INSPstI x INSDraIII es el que presenta una mayor intensidad de asociación (0,985) frente al par H19Alul x TpHDraII que es el que presenta la menor intensidad de asociación (0,130), medida en términos de D' .

Tabla 24. DG significativos entre pares de SNPs

Par de SNPs	D'	sin Yates		con Yates	
		χ^2	P	χ^2	P
INSPstI x INSDraIII	0,985	357,47	0,000	356,09	0,000
INSPstI x H19Alul	0,229	5,89	0,015	5,75	0,017
INSDraIII x H19Alul	0,262	8,50	0,004	8,33	0,004
INSDraIII x H19Rsal	-0,234	4,71	0,030	4,58	0,032
DRD4Smal x HRASNsil	-0,651	8,39	0,004	8,11	0,004
H19Alul x H19Rsal	-0,982	271,62	0,000	270,79	0,000
H19Alul x TpHDraII	0,130	6,03	0,014	5,91	0,015
H19Rsal x TpHDraII	-0,150	5,56	0,018	5,45	0,020
RRM1TaqI x ART1Hpall	0,219	8,05	0,005	7,88	0,005
TSC0683096 x TSC0683097	0,961	319,03	0,000	317,94	0,000

En la figura 10 se muestra la distribución de las frecuencias de los valores de intensidad, en términos de D' , de los DG significativos. En la tabla 25 se

Resultados

muestran los estadísticos descriptivos de la distribución de los valores de D' significativos, así como su ajuste a la distribución normal.

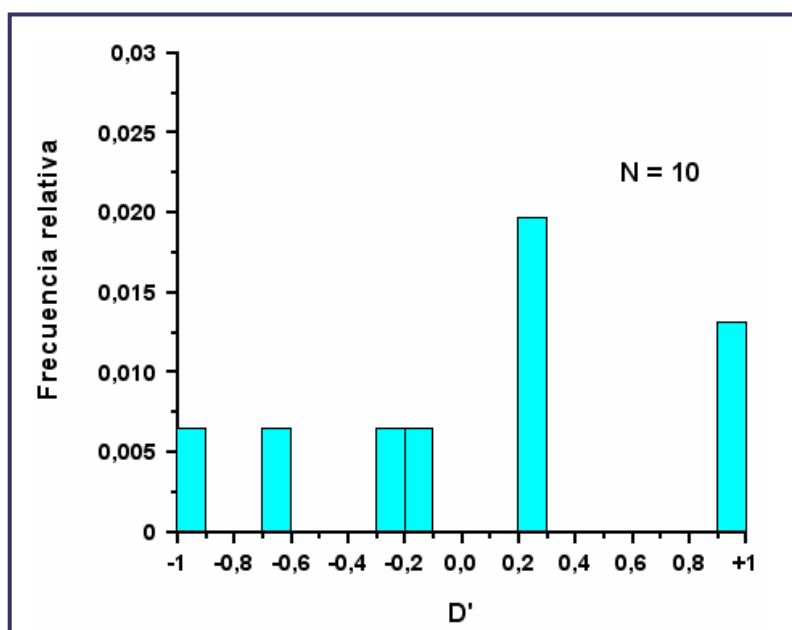


Figura 10. Distribución de frecuencias de la intensidad (D') de los desequilibrios de los pares de loci significativos. N es el número de pares de loci en desequilibrio significativo.

Tabla 25. Estadísticos descriptivos, valores de sesgo y curtosis y ajuste a la distribución normal ($D_{(KS)}$) de las distribuciones de los D' significativos. N indica el número de casos significativos.

	D'	$ D' $
N	10	10
%	6,5 % (10/153)	6,5 % (10/153)
Rango	-0,982 – 0,985	0,130 – 0,985
Media $\pm$ SE	0,077 $\pm$ 0,197	0,480 $\pm$ 0,117
Varianza	0,39	0,14
Sesgo (g_1)	-0,09 $\pm$ 0,69	
Curtosis (g_2)	-0,20 $\pm$ 1,33	
$D_{(KS)}$	0,183*	

*P < 0,05

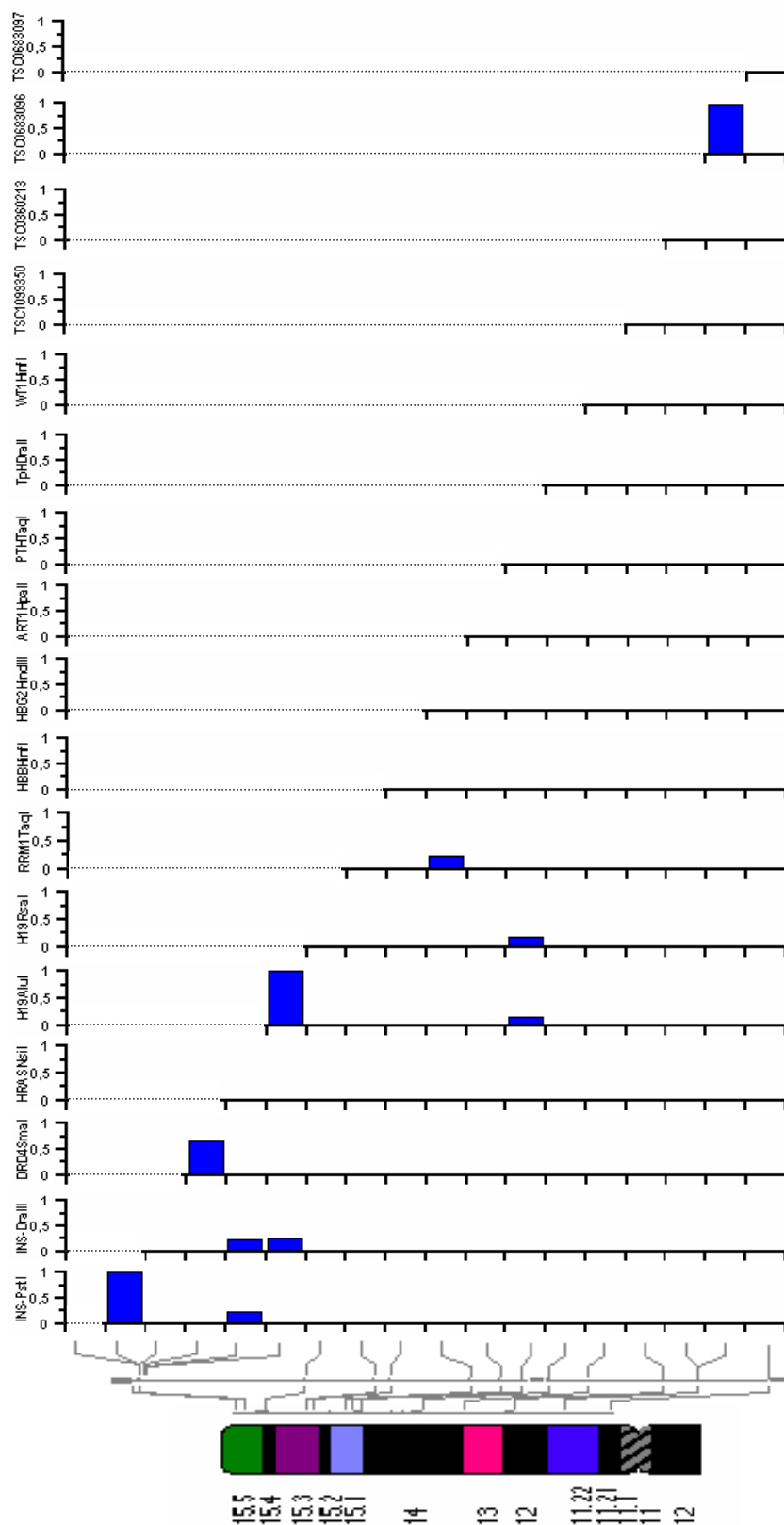
Se observa que los valores de D' abarcan casi todo el rango posible de intensidades con un mínimo de -0,982, un máximo de 0,985 y un promedio ($\pm$ SE) de $0,077 \pm 0,197$. El rango de la intensidad de los DG en valor absoluto presenta un mínimo de 0,130 y un máximo de 0,985. Así, el promedio ($\pm$ SE) de los valores de D' en valor absoluto resultó en un valor de intensidad moderada $0,480 \pm 0,117$. El test de Kolmogorov-Smirnov muestra que la distribución de los valores de D' significativos observados se desvía significativamente de la normalidad ($P < 0,05$). El sesgo presenta un valor ligeramente negativo debido a la asimetría de la distribución, con un mayor número de casos de signo positivo. La curtosis tiene un valor negativo lo que indica un exceso de casos en los extremos de la distribución.

1.3.2. Relación del DG con la FRE y la DF

La distribución del DG entre pares de SNPs a lo largo del brazo corto del cromosoma 11 se muestra en la figura 11. La figura muestra que la distribución del DG entre pares de SNPs no es aleatoria en el cromosoma, sino que parecen darse más frecuentemente asociaciones significativas entre los pares de loci más cercanos.

(siguiente página) Figura 11. Distribución del DG entre pares de SNPs a lo largo del brazo corto del cromosoma 11. Se muestra en abscisas el ideograma de 11p donde se señala la posición relativa de los SNPs desde el telómero (INSPstI) al centrómero (TSC0683097). Y en ordenadas se muestra la intensidad del DG para los pares de SNPs, medido en términos de D' .

Resultados



En la tabla 26 se exponen los valores descriptivos de la FRE y DF entre los pares de SNPs en DG significativo. Los desequilibrios se extienden por buena parte de la región analizada con valores máximos de FRE de 19,98 cM (H19Alul x TphDrall y H19Rsal x TphDrall) y mínimos de 0 cM (INSPstI x INSDraIII, H19Alul x H19Rsal y TSC0683096 x TSC0683097). Sin embargo, en la mayoría de los pares de SNPs significativos (70 %) la FRE es inferior a 1 cM. De hecho, la FRE promedio ($\pm$ SE) entre los loci en DG significativo fue de $5,1 \pm 2,6$ cM. Incluso este valor es elevado por el efecto que ejercen los dos pares de SNPs en DG significativo con una FRE de ~ 20 cM. En cuanto a la DF, la distancia máxima a la que se detectó DG significativo fue de 16,1 Mb (H19Alul x TphDrall, H19Rsal x TphDrall) y la distancia mínima de ~ 0 Mb (INSPstI x INSDraIII, H19Alul x H19Rsal y TSC0683096 x TSC0683097). El promedio de la DF entre los pares de SNPs en DG significativo fue $3,3 \pm 2,1$ Mb.

Tabla 26. Frecuencia de recombinación efectiva (cM) y distancia física (Mb) para pares de SNPs en desequilibrio significativo

	cM	Mb
Mínimo	0	1×10^{-4}
Máximo	20,0	16,1
Media $\pm$ SE	$5,1 \pm 2,6$	$3,3 \pm 2,1$

En la tabla 27 se encuentran representados los valores de FRE entre todos los pares de SNPs analizados. En sombreado se destacan los pares de loci en desequilibrio significativo. El código de colores indica la intensidad de desequilibrio que presentan esos pares de loci. Parece que los desequilibrios no se encuentran distribuidos aleatoriamente por la región analizada sino que aparecen mayoritariamente en aquellos pares con menor FRE. De hecho, de los 10 pares de SNPs en DG significativo, el 70 % (7/10) están separados por menos de 1 cM. Así mismo, atendiendo a la intensidad del DG entre los pares de SNPs separados por menos de 1 cM, el 57 % (4/7) presentan una intensidad de DG, medida en

Resultados

términos de D' , mayor de 0,5. Así, parece que la FRE entre pares de SNPs está influyendo tanto en la cantidad como en la intensidad de las asociaciones.

Tabla 27. Frecuencias de recombinación efectivas (cM) entre pares de SNPs. En sombreado se destacan los pares de loci en DG significativo

LOCUS	INS-PstI	INS-DraIII	DRD4-Smal	HRAS-Nsil	H19-AluI	H19-RsaI	RRM1-TaqI	HBB-HinfI	HBB2-HindIII	ARTI-HpaI	PTH-TaqI	TpH-DraII	WT1-HinfI	TSC1099350	TSC0360213	TSC0683096	TSC0683097
INS-DraIII	0,00																
DRD4-Smal	0,01	0,01															
HRAS-Nsil	0,44	0,44	0,43														
H19-AluI	0,45	0,45	0,44	0,01													
H19-RsaI	0,45	0,45	0,44	0,01	0,00												
RRM1-TaqI	9,19	9,19	9,18	8,77	8,76	8,76											
HBB-HinfI	12,77	12,77	12,76	12,36	12,35	12,35	3,80										
HBB2-HindIII	13,54	13,54	13,54	13,14	13,13	13,13	4,61	0,81									
ARTI-HpaI	17,77	17,77	17,77	17,39	17,38	17,38	9,16	5,43	4,63								
PTH-TaqI	19,27	19,27	19,26	18,90	18,89	18,89	10,83	7,14	6,35	1,74							
TpH-DraII	20,36	20,36	20,35	19,99	19,98	19,98	12,05	8,39	7,61	3,03	1,29						
WT1-HinfI	30,94	30,94	30,93	30,67	30,66	30,66	24,58	21,58	20,93	16,93	15,37	14,19					
TSC1099350	31,76	31,76	31,75	31,50	31,49	31,49	25,70	22,82	22,16	18,20	16,69	15,54	2,08				
TSC0360213	34,28	34,28	34,27	34,05	34,04	34,04	28,78	26,13	25,54	21,90	20,48	19,39	5,70	4,34			
TSC0683096	35,16	35,16	35,16	34,95	34,94	34,94	30,00	27,50	26,91	23,34	21,98	20,93	7,76	6,27	1,95		
TSC0683097	35,16	35,16	35,16	34,95	34,94	34,94	30,00	27,50	26,91	23,34	21,98	20,93	7,76	6,27	1,95	0,00	
TSC1204767	37,04	37,04	37,04	36,86	36,85	36,85	32,58	30,42	29,84	26,50	25,28	24,31	12,11	10,75	6,53	4,66	4,66

$D' = 0 \quad 0,25 \quad 0,50 \quad 0,75 \quad 1,00$

Con el objeto de alcanzar un apoyo estadístico a las observaciones anteriores, se realizó un test χ^2 en tablas de contingencia 2 x 2 para comprobar la existencia de un mayor número de asociaciones no al azar entre loci próximos que entre loci alejados. Para este fin la muestra se dividió en dos grupos en función del grado de ligamiento de los SNPs implicados en cada haplotipo. Así, por loci próximos entenderemos los que estén separados por distancias ≤ 2 cM o 2 Mb y por loci alejados los separados por más de 2 cM o 2 Mb. Esta división entre loci próximos y loci alejados ya ha sido utilizada en otras ocasiones para abordar la problemática de la relación entre el DG y la FRE o DF (PETERSON *et al.* 1995; LAAN y PÄÄBO 1997). La tabla 28 muestra que la comparación resulta altamente

significativa, tanto cuando se separan los loci en función de la FRE como de la DF ($P < 0,01$).

Tabla 28. Test χ^2 junto a su probabilidad (P), de la comparación del número de asociaciones significativas y no significativas entre loci estrecha y débilmente ligados, según la FRE (cM) y la DF (Mb)				
	≤ 2 cM	> 2 cM	≤ 2 Mb	> 2 Mb
Significativos	7	3	8	2
No significativos	13	130	22	121
	$\chi^2 = 30,4$; $P = 0,000000$		$\chi^2 = 24,8$; $P = 0,000001$	

Un aspecto importante a elucidar es comprobar si existe una tendencia monótona hacia una menor intensidad del DG con una mayor FRE o DF entre los pares de SNPs. Con este fin se realizó un análisis de correlación entre la intensidad del desequilibrio y la FRE o la DF. Se empleó para ello el coeficiente de correlación paramétrico r de Pearson y el coeficiente de correlación no paramétrico de rangos τ de Kendall, la probabilidad se obtuvo aplicando el test de Mantel. El estudio detectó que tanto con la FRE como con la DF la correlación es negativa y altamente significativa ($P < 0,01$), como cabría esperar en base a la teoría genético poblacional (tabla 29). Por tanto, al menos en el brazo corto del cromosoma 11 parece que la recombinación ejerce un efecto predominante sobre otros factores evolutivos al homogeneizar las frecuencias gaméticas de loci lo suficientemente separados en el cromosoma.

Los resultados anteriores apoyan estadísticamente la existencia de una relación entre la cantidad e intensidad del DG con la FRE y la DF. Además, estos resultados sugieren que los desequilibrios significativos detectados entre pares de SNPs no son espúreos. Si el DG hubiese sido generado sólo por azar o error de muestreo, los DG estarían distribuidos aleatoriamente a lo largo de la región analizada. Por el contrario, los desequilibrios parecen estar distribuidos a lo largo de 11p en función de la FRE y la DF.

Resultados

Tabla 29. Correlación de los D' significativos con la frecuencia de recombinación (FRE) y distancia física (DF). Se indica el coeficiente r de Pearson, el coeficiente de rangos τ de Kendall y la probabilidad (P) calculada mediante el test de Mantel. N indica el número de asociaciones no al azar significativas.

	N	r	τ	P
FRE	10	-0,584	-0,919	0,008
DF	10	-0,497	-0,822	0,009

La figura 12 muestra gráficamente la correlación de la intensidad del DG, medida en términos de D' , con la FRE y la DF. Se observa una tendencia hacia una menor intensidad de las asociaciones al aumentar la distancia entre los SNPs. Esta representación gráfica sugiere que la relación entre la intensidad del DG y la FRE o DF entre los pares de SNPs no es monótona. La correlación significativa se mantiene por el efecto que ejercen en la correlación los cuatro pares de SNPs estrechamente ligados que presentan un valor de D' elevado ($> 0,5$). Así, el valor de τ es más elevado que r , lo que denota que la relación no sigue una tendencia lineal, sino curvilínea.

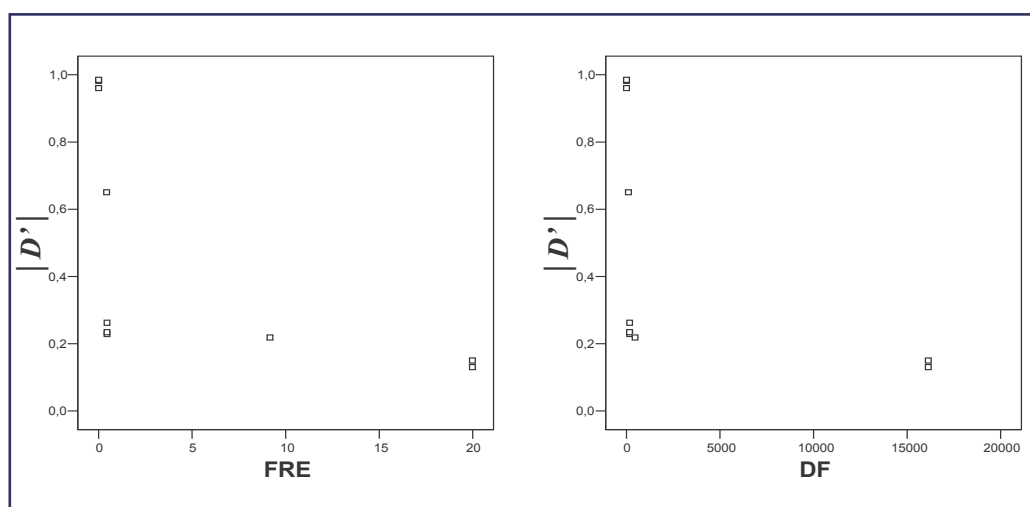


Figura 12. Gráficas de la correlación de la intensidad del DG ($|D'|$) con la FRE y la DF.

2. ESTUDIO DEL DG ENTRE PARES DE LOCI MICROSATÉLITES

2.1. Análisis de la variabilidad y estructura genética poblacional

El análisis de la estructura genética poblacional, además de tener interés por sí misma, será de gran ayuda a la hora de interpretar los resultados en los patrones del DG a lo largo de 11p. Por otro lado, el cumplimiento de las proporciones Hardy-Weinberg nos asegurará una estimación fiable de las frecuencias haplotípicas aplicando el algoritmo EM.

En la tabla 30 se muestra el número de alelos detectados en la muestra y su frecuencia relativa para cada uno de los 22 loci microsatélites analizados, así como el tamaño de cada uno de ellos en pb. El número de alelos por locus fue en general alto y osciló entre 5 (D11S1330) y 16 (D11S4080), con un valor medio de $11 \pm 0,6$ alelos por locus. Este número de alelos por locus ha sido ligeramente superior al previamente encontrado para esos loci (DIB *et al.* 1996), probablemente debido al mayor tamaño de muestra analizado en el presente estudio. Las frecuencias alélicas relativas oscilaron entre 0,001 y 0,852 con un valor medio de $0,094 \pm 0,009$. Se puede observar que en casi todos los loci aparecen alelos que se detectan una sola vez en la muestra (en cursiva en la tabla 30). La frecuencia de estos alelos de copia única o *singletons* para el conjunto de los loci fue del 13 % (31/235). La presencia de estos *singletons* podría ser debida a la elevada tasa mutacional que poseen los microsatélites (DI RIENZO *et al.* 1998). Estos *singletons* se eliminaron del análisis del DG por las razones que se detallarán posteriormente. Una vez eliminados los *singletons*, las frecuencias alélicas presentaron un rango entre 0,003 y 0,852 con una frecuencia promedio para todos los loci de $0,108 \pm 0,010$.

Tabla 30. Frecuencias alélicas de los loci microsatélites

Alelo	F. Rel.	Alelo	F. Rel.	Alelo	F. Rel.	Alelo	F. Rel.
D11S4177		D11S4124		D11S1760		D11S909	
177	0,0012	163	0,0025	72	0,0025	95	0,0012
183	0,3086	165	0,0099	80	0,1346	109	0,0111
185	0,0012	167	0,0519	82	0,2568	113	0,0136
199	0,0037	169	0,3531	84	0,0407	115	0,0407
201	0,1185	171	0,3667	86	0,0161	117	0,1753
203	0,0346	173	0,1642	90	0,3383	119	0,4272
205	0,0877	175	0,0111	92	0,0124	121	0,1815
207	0,0728	177	0,0370	94	0,1185	123	0,1173
209	0,2938	181	0,0037	96	0,0025	125	0,0247
211	0,0444	D11S4181		98	0,0679	127	0,0049
213	0,0296	193	0,0037	100	0,0062	129	0,0012
215	0,0012	201	0,0025	108	0,0037	131	0,0012
217	0,0025	203	0,0309	D11S1331		D11S4121	
D11S1318		205	0,0198	183	0,0012	92	0,1235
121	0,0012	207	0,1926	191	0,0383	96	0,0025
123	0,0827	209	0,1198	193	0,0790	104	0,0111
125	0,0012	211	0,0605	195	0,2901	106	0,0543
127	0,0864	213	0,3519	197	0,3593	108	0,1988
129	0,2914	215	0,1420	199	0,0407	110	0,2580
131	0,0741	217	0,0630	201	0,1407	112	0,1370
133	0,0235	219	0,0099	203	0,0062	114	0,0284
135	0,0037	221	0,0025	205	0,0136	116	0,0124
137	0,0506	231	0,0012	207	0,0025	118	0,0370
139	0,2889	D11S1323		209	0,0012	120	0,0457
141	0,0667	197	0,0012	211	0,0037	122	0,0852
143	0,0235	199	0,0012	213	0,0086	124	0,0062
145	0,0062	201	0,0383	215	0,0111	D11S926	
D11S4146		203	0,5691	217	0,0037	131	0,0111
193	0,0012	205	0,1210	D11S4188		133	0,0049
195	0,0617	207	0,1988	101	0,0012	135	0,1222
197	0,4370	209	0,0617	109	0,0012	137	0,4568
199	0,0358	211	0,0074	111	0,0765	139	0,1309
201	0,0062	213	0,0012	113	0,0049	141	0,2148
203	0,0272			115	0,3914	143	0,0506
205	0,2691			117	0,1901	145	0,0086
207	0,0469			119	0,1049		
209	0,0951			121	0,0951		
211	0,0099			123	0,1136		
213	0,0086			125	0,0161		
219	0,0012			127	0,0025		
				131	0,0012		
				133	0,0012		

Tabla 30. (continuación)

Alelo	Frec. Relat.	Alelo	Frec. Relat.	Alelo	Frec. Relat.	Alelo	Frec. Relat.
D11S4080		D11S4185		D11S1330		D11S903	
116	0,0012	137	0,0074	154	0,0012	93	0,0012
128	0,0012	139	0,0222	156	0,8519	97	0,0062
130	0,0049	141	0,5049	158	0,0580	99	0,3840
132	0,0037	143	0,4173	160	0,0840	101	0,0951
134	0,0222	145	0,0420	162	0,0049	103	0,1926
136	0,0469	147	0,0049	D11S1360		105	0,0914
138	0,3753	149	0,0012	103	0,0198	107	0,1852
140	0,1210	D11S935		105	0,0099	109	0,0407
142	0,1667	188	0,0049	107	0,0346	111	0,0037
144	0,0407	192	0,0012	109	0,4926	D11S1361	
146	0,0790	196	0,3519	111	0,3000	202	0,0025
148	0,0605	198	0,0284	113	0,1012	206	0,6654
150	0,0444	200	0,0741	115	0,0321	208	0,1642
152	0,0247	202	0,0309	117	0,0074	210	0,0198
154	0,0062	204	0,1099	119	0,0012	212	0,0074
158	0,0012	206	0,3741	121	0,0012	216	0,0457
D11S1324		208	0,0247	D11S1763		218	0,0161
110	0,1556	D11S4148		182	0,0136	220	0,0654
112	0,0124	170	0,0074	184	0,0136	222	0,0124
114	0,0037	172	0,0086	186	0,4432	224	0,0012
116	0,0086	174	0,0111	188	0,4210		
118	0,0864	176	0,2667	190	0,1049		
120	0,1778	178	0,3593	194	0,0037		
122	0,2296	180	0,2815				
124	0,1679	182	0,0494				
126	0,0728	184	0,0111				
128	0,0691	186	0,0037				
130	0,0161	188	0,0012				

En la figura 13 se encuentran representadas las distribuciones de las frecuencias de los tamaños de cada alelo por locus. El tamaño de los alelos cubre un amplio rango con un valor mínimo de 72 pb (D11S1760), un valor máximo de 231 pb (D11S4181) y un tamaño medio para el conjunto de los loci de $157 \pm 2,8$ pb. Ninguna distribución se ajusta a la normalidad según el test de Kolmogorov-Smirnov ($P < 0,01$). Se puede observar que en algunos casos las distribuciones presentan un carácter bimodal debido a la existencia de dos grupos de alelos

Resultados

diferenciables por su tamaño y separados por clases alélicas de tamaño intermedio ausentes en la muestra o en muy baja frecuencia. Este tipo de distribución se observa más claramente en el locus D11S4177, donde la clase alélica más frecuente (alelo 183) se encuentra separada del resto por varias clases alélicas de tamaño intermedio ausentes en la muestra. Se ha postulado que, en algunos casos, las distribuciones bimodales podrían estar causadas por dos subgrupos de alelos con alguna diferencia en su secuencia, de manera que los alelos de un grupo no podrían mutar a alelos del otro grupo y viceversa (BRINKMAN *et al.* 1998). Cualquier distribución alélica es el resultado de un proceso mutacional que genera nuevos alelos, junto con la deriva genética y la selección modelando las frecuencias alélicas. Además, factores de la historia poblacional como cuellos de botella, rápido crecimiento poblacional y migración podrían jugar un importante papel (SHRIVER *et al.* 1993; VALDES *et al.* 1993; STEPHAN y KIM 1998; SCHLÖTTERER 2000; CALABRESE *et al.* 2001; HUANG *et al.* 2002; ELLEGREN 2004; XU y FU 2004).

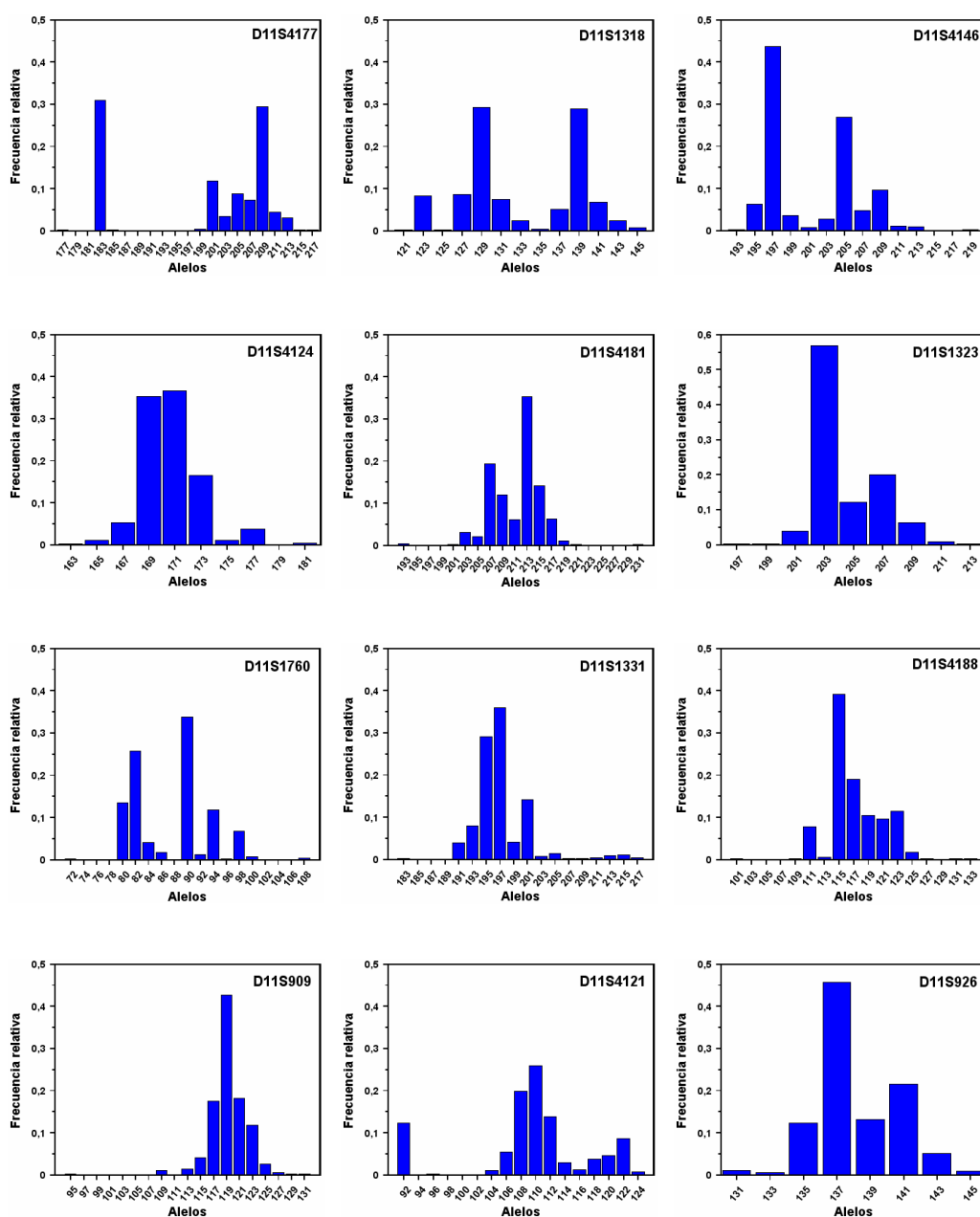


Figura 13. Distribuciones de las frecuencias alélicas de los 22 microsatélites analizados

Resultados

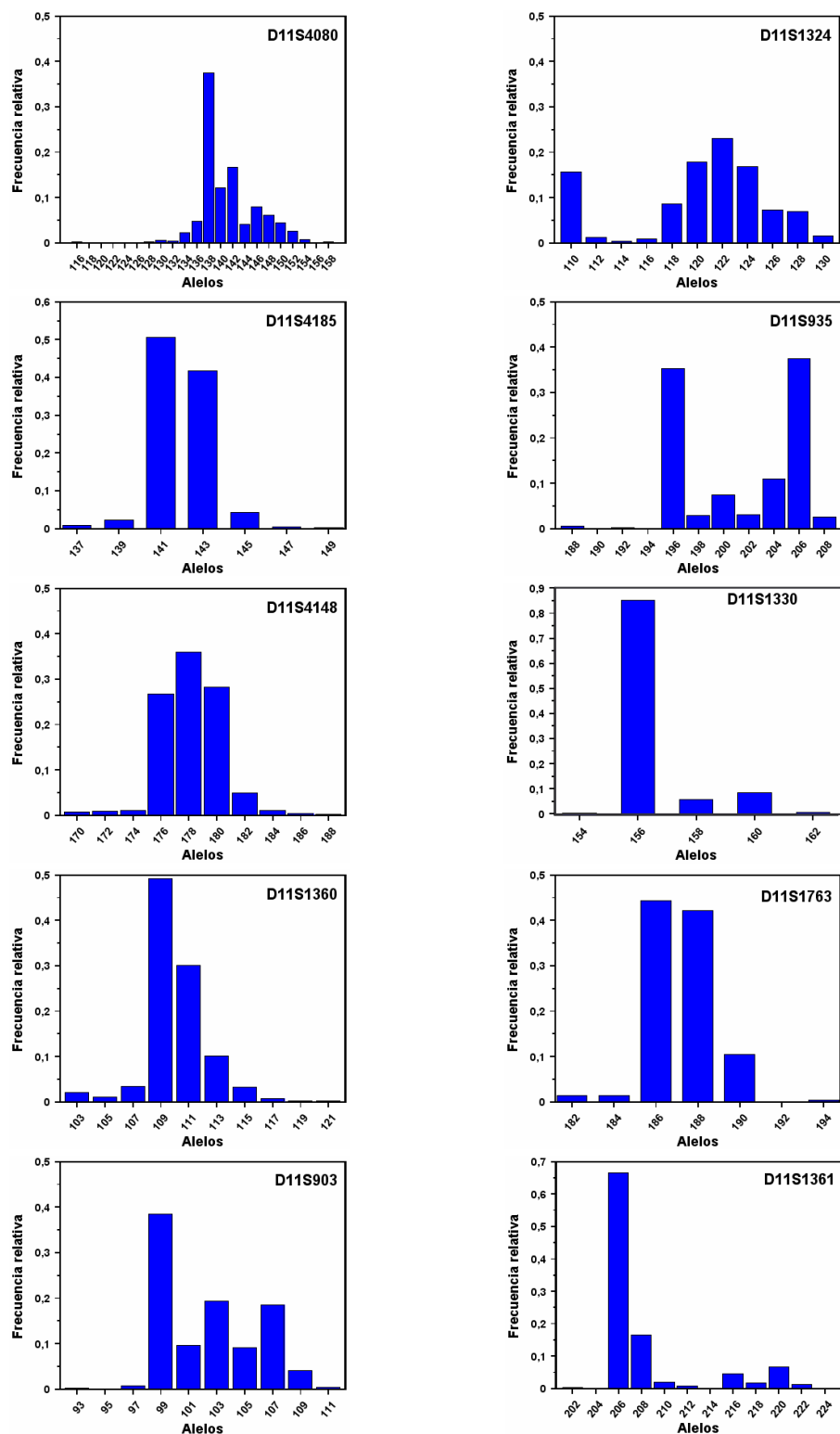


Figura 13. (continuación)

En la tabla 31 se muestran las estimas de variabilidad para cada microsatélite analizado así como la magnitud y significación de las desviaciones de las proporciones H-W por locus. Los microsatélites poseen un alto nivel de variabilidad, como cabría esperar al estar ante loci multialélicos. La heterozigosis observada varía tremendamente entre los loci con un mínimo de heterozigosis observada de 0,26 (D11S1331) y un máximo de 0,84 (D11S4121). La media de la heterozigosis observada para el conjunto de los loci resultó bastante elevada con un valor de $0,71 \pm 0,03$. Este nivel de heterozigosis ha sido observado en otras ocasiones en diferentes estudios con microsatélites (DEKA *et al.* 1995; DIB *et al.* 1996). Las heterozigosis observada y esperada para cada uno de los loci poseen una buena concordancia. De hecho, los promedios para el conjunto de los loci presentan el mismo valor. Las desviaciones de las proporciones H-W medidas mediante el coeficiente f_T (ROBERTSON y HILL 1984) fueron en general bajas y oscilaron (en valor absoluto) entre $0,001 \pm 0,018$ (D11S4124) y $0,061 \pm 0,048$ (D11S1331), con un valor medio de $0,014 \pm 0,003$. Estas desviaciones no son significativas en ningún caso excepto en los loci D11S1331 y D11S4121 ($P < 0,01$). En ambos microsatélites las desviaciones se deben a la presencia de alelos en muy baja frecuencia, inferior al 1% (alelos 213 y 217 en D11S1331 y alelo 124 en D11S4121 en la tabla 29). El estudio de las desviaciones de las proporciones H-W por alelo en términos del coeficiente f_{ii} (ROBERTSON y HILL 1984), mostró que en estos tres casos las desviaciones son muy elevadas, positivas y significativas (0,668, 0,281 y 0,398 respectivamente); véase HERMIDA (2001); ZAPATA *et al.* (2001b). El exceso significativo de homocigotos detectado en ambos loci está ocasionado por la existencia de un individuo homocigoto para cada uno de los alelos mencionados, siendo sumamente bajo el valor esperado de esos genotipos. Por tanto, las desviaciones significativas de las proporciones H-W en los loci D11S1331 y D11S4121 están ocasionadas por las limitaciones estadísticas del test χ^2 cuando los valores esperados son muy bajos (LESSIOS 1992). Al examinar las desviaciones por alelo en todos los loci en términos de f_{ii} la magnitud de las desviaciones es baja, en general, con un rango de 0 a 0,668 (alelo 213 de D11S1331) y una media de $0,032 \pm 0,004$. Sólo un 3 % (7/235), presentaron desviaciones significativas de las proporciones H-W. Tres de ellos son los

Resultados

anteriormente mencionados y los cuatro restantes (alelo 183 D11S4177, alelo 125 D11S909, alelo 198 D11S935 y alelo 182 D11S1763) también presentan un exceso considerable de homocigotos debido igualmente a la sensibilidad del test χ^2 cuando los valores esperados son muy bajos. Globalmente la proporción de alelos que se desvían de las proporciones H-W es baja (3 %). Así, estas desviaciones podrían ser debidas a un error de tipo I al estar por debajo del nivel de significación nominal ($\alpha = 0,05$).

Tabla 31: Magnitud ($\hat{f}_T$) y significación (χ^2) de las desviaciones de las proporciones Hardy-Weinberg para cada uno de los microsatélites analizados.

Locus	Heterozigosis observada	Heterozigosis esperada	$\hat{f}_T \pm SE$	P
D11S4177	0,795	0,788	0,005 $\pm$ 0,020	0,734
D11S1318	0,817	0,805	-0,001 $\pm$ 0,011	0,943
D11S4146	0,736	0,720	-0,004 $\pm$ 0,002	0,793
D11S4124	0,723	0,710	0,001 $\pm$ 0,018	0,946
D11S4181	0,770	0,797	0,012 $\pm$ 0,024	0,414
D11S1323	0,625	0,617	0,001 $\pm$ 0,018	0,941
D11S1760	0,807	0,782	-0,015 $\pm$ 0,000	0,310
D11S1331	0,775	0,758	0,061 $\pm$ 0,048	0,000
D11S4188	0,765	0,773	0,009 $\pm$ 0,027	0,519
D11S909	0,723	0,738	0,024 $\pm$ 0,039	0,113
D11S4121	0,837	0,846	0,045 $\pm$ 0,028	0,002
D11S926	0,696	0,711	-0,002 $\pm$ 0,017	0,914
D11S4080	0,795	0,801	0,005 $\pm$ 0,018	0,678
D11S1324	0,832	0,846	0,001 $\pm$ 0,015	0,976
D11S4185	0,533	0,569	0,020 $\pm$ 0,044	0,328
D11S935	0,716	0,717	0,026 $\pm$ 0,036	0,145
D11S4148	0,716	0,719	-0,004 $\pm$ 0,007	0,829
D11S1330	0,262	0,264	-0,006 $\pm$ 0,000	0,822
D11S1360	0,672	0,655	-0,006 $\pm$ 0,000	0,728
D11S1763	0,612	0,616	0,037 $\pm$ 0,043	0,095
D11S903	0,775	0,763	-0,007 $\pm$ 0,000	0,700
D11S1361	0,531	0,524	-0,013 $\pm$ 0,000	0,419
MEDIA $\pm$ S.E.	0,705 $\pm$ 0,028	0,705 $\pm$ 0,028	0,014 $\pm$ 0,003	

La media de $\hat{f}_T$ está calculada a partir de los valores absolutos

P es la probabilidad obtenida mediante el test chi-cuadrado.

2.2. ANÁLISIS DEL DESEQUILIBRIO GAMÉTICO

Un total de 20.706 haplotipos de dos loci se pueden evaluar cuando consideramos sólo los alelos presentes más de una vez en la muestra (204 alelos del total de 235). El hecho de eliminar del análisis estos *singletons* o alelos de copia única deriva de su influencia en la estimación del desequilibrio. Así, los haplotipos relacionados con estos alelos podrían provocar una sobreestima tanto en la intensidad como en la significación del desequilibrio al estar ausentes de la muestra algunas clases haplotípicas (ZAPATA & ALVAREZ 1992). La contribución de los alelos en baja frecuencia al desequilibrio de la población, se infirió repitiendo el análisis del DG para sólo los alelos con una frecuencia mayor del 3 %. Estos alelos en baja frecuencia podrían ser de reciente incorporación en la población debido a la elevada tasa de mutación que poseen los microsatélites (entre $2,5 \times 10^{-5}$ y 10^{-2} eventos mutacionales por generación; WEBER y WONG 1993) y al tiempo que se requiere para que un alelo nuevo en la población aumente su frecuencia.

El análisis del DG entre pares de loci se ha realizado a dos niveles: Un análisis de los DG interalélicos y un análisis global del DG resumiendo la información contenida en los DG interalélicos. La estimación del desequilibrio interalélico y global se realizó en función del signo. Tras el cómputo de todos los desequilibrios interalélicos se incluyeron solamente aquellos de signo positivo con el fin de obtener unos niveles más precisos de las asociaciones existentes en el cromosoma 11p. Esta decisión viene motivada por la observación de que la potencia estadística del test chi-cuadrado es mayor para detectar desequilibrios positivos que negativos, cuando las clases haplotípicas en acoplamiento contienen los alelos más o menos comunes. Por otro lado, la varianza de muestreo de las asociaciones negativas es mayor que las positivas cuando las frecuencias alélicas no son intermedias, por lo que podrían producir una sobreestima de la intensidad del desequilibrio (véase *Material y Métodos* y ZAPATA *et al.* 2001a, b). De esta manera, si considerásemos todo el conjunto de valores de los desequilibrios interalélicos y globales (negativos y positivos), estaríamos introduciendo un sesgo

en nuestros resultados y una infraestima en la proporción de desequilibrio significativo existente en la población de estudio.

2.2.1. DG interalélicos

El cómputo de los DG interalélicos de signo positivo totales resultó en una cantidad de 7512, de los cuales 3417 implicaban a alelos con frecuencia mayor al 3 %. La significación del DG interalélico positivo se obtuvo a partir de las frecuencias haplotípicas, mediante tests χ^2 con la corrección de Yates en tablas de contingencia 2 x 2. El porcentaje de DG interalélicos positivos significativos fue del 16% (1211/7512) considerando todos los alelos y del 13% (449/3417) considerando sólo los alelos más frecuentes ($p_i, q_j > 3\%$).

La figura 14 muestra la distribución de los porcentajes de los DG interalélicos positivos significativos por par de loci. El 98 % (227/231) de los pares de loci muestran algún DG interalélico positivo significativo considerando la muestra de todos los alelos. Sin embargo, cuando se consideran sólo los alelos más frecuentes, el porcentaje de pares de loci con algún DG interalélico positivo significativo, desciende al 79 % (183/231). Los porcentajes de los DG interalélicos positivos no estuvieron distribuidos uniformemente a lo largo de los pares de loci. El porcentaje de asociaciones interalélicas positivas por par de loci varió entre el 4% (D11S1324 x D11S1763, D11S909 x D11S1360, D11S1760 x D11S1763) y el 50% (D11S4185 x D11S935), siendo la media $16,3 \pm 0,5$. Entre los alelos más frecuentes estas cifras no difieren mucho. Así, los porcentajes variaron entre un 4,2% (D11S4181 x D11S4188) y un 50% (D11S4124 x D11S1361), con un valor promedio de $16,3 \pm 0,6$.

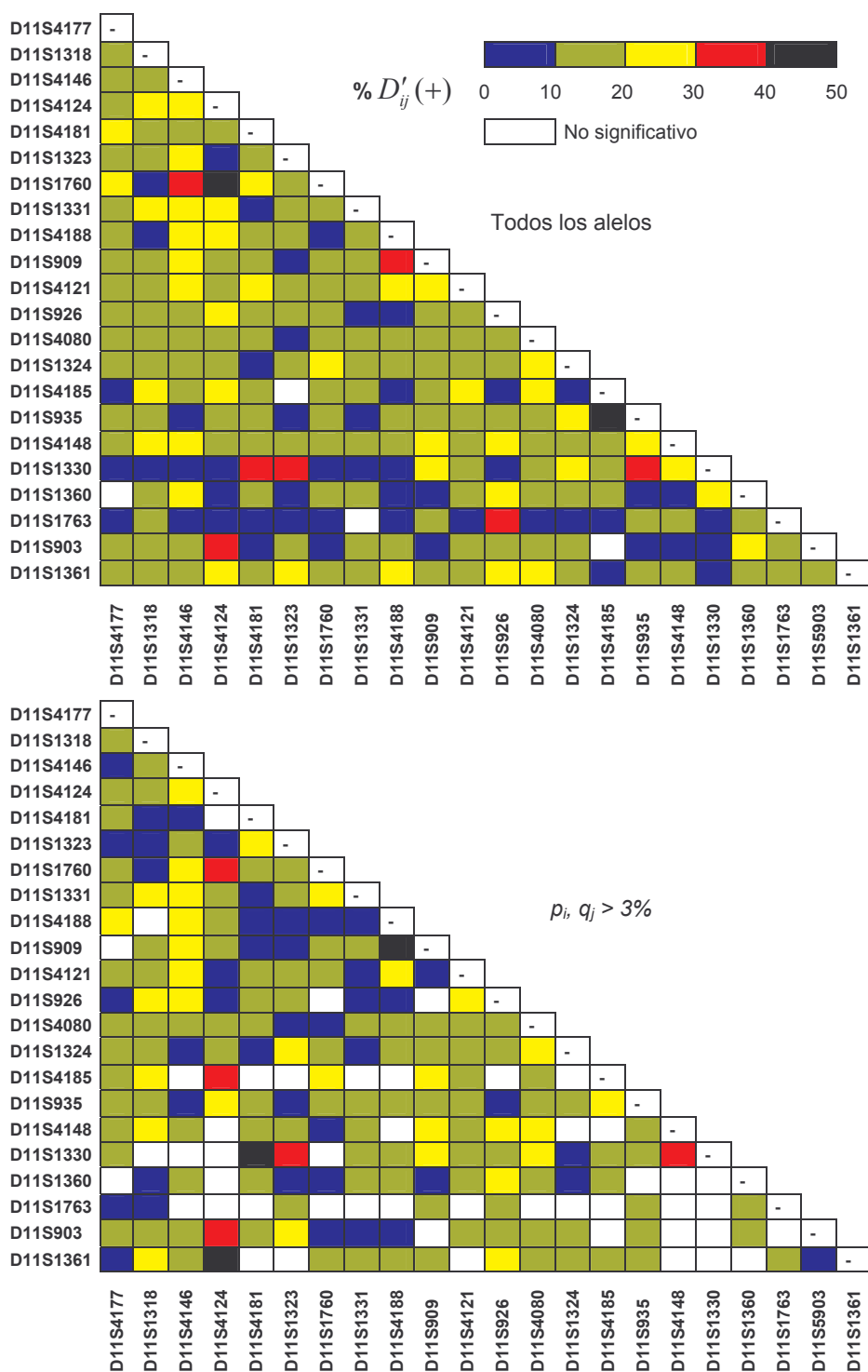


Figura 14. Porcentaje de $D'_{ij}(+)$ significativos para cada par de microsatélites.

Resultados

La figura 15 muestra las distribuciones de los valores de intensidad, medida en términos de $D'_{ij}(+)$, de los DG interalélicos positivos significativos para todos los alelos y para los alelos más frecuentes. Se observa la gran cantidad de disequilibrios de intensidad moderada que están presentes tanto en la muestra de todos los alelos como en la muestra de los alelos más frecuentes. Es de destacar que los disequilibrios de mayor intensidad ($D'_{ij}(+) > 0,6$) únicamente se detectan cuando se consideran sólo los alelos más frecuentes.

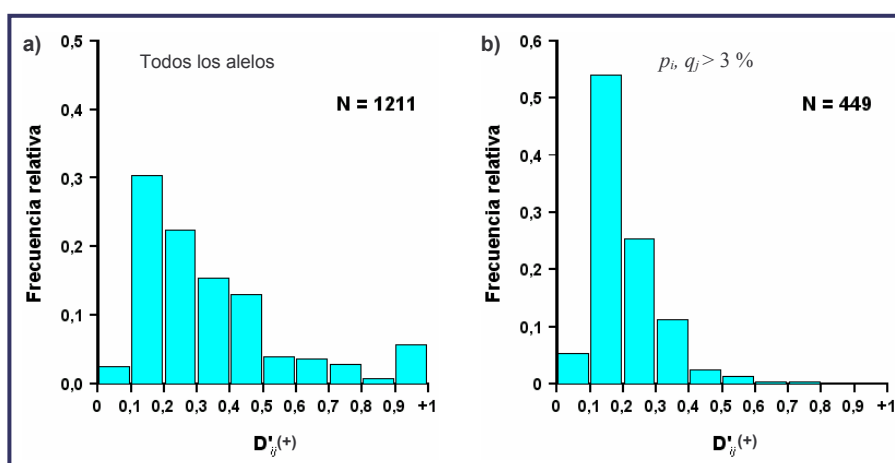


Figura 15. Distribución de las frecuencias de los valores de intensidad del DG interalélico positivo entre pares de alelos de microsatélites. a) Todos los alelos. b) Alelos con frecuencia mayor al 3%. N es el número de pares de alelos en disequilibrio significativo.

En la tabla 32 se muestran los estadísticos descriptivos de la distribución de los valores de $D'_{ij}(+)$ para todos los alelos y para los alelos más frecuentes. Estos valores cubren un amplio rango y en el caso de todos los alelos van desde valores muy bajos (0,09) hasta el máximo valor que el coeficiente $D'_{ij}(+)$ puede tomar (1,00). A pesar de estar presentes valores máximos de intensidad, la media refleja una intensidad moderada ($0,34 \pm 0,01$). A modo ilustrativo, 853 casos del

total de 1211 significativos (70%), presentan valores de $D'_{ij}(+)$ menores de 0,40. Es de destacar que al eliminar los alelos con frecuencia menor al 3% la intensidad máxima detectada es claramente menor que al considerar todos los alelos (0,77). Así mismo, el promedio de intensidad es más bajo con los alelos más frecuentes ($0,21 \pm 0,01$).

Tabla 32. Estadísticos descriptivos de las distribuciones de los $D'_{ij}(+)$ significativos. N indica el número de pares de alelos en desequilibrio.

	Todos los alelos	Alelos de frecuencia mayor al 3%
N	1211	449
%	16% (1211/7512)	13 % (449/3417)
Rango	0,09 – 1,00	0,09 – 0,77
Media $\pm$ SE	$0,34 \pm 0,01$	$0,21 \pm 0,01$
Varianza	0,05	0,01

La figura 16 muestra la distribución de los valores promedio de los $D'_{ij}(+)$ significativos [$\overline{D'_{ij}(+)}$] por par de loci. Los valores oscilaron entre 0,10 (D11S4177 x D11S1330) y 0,89 (D11S4148 x D11S1763), con un valor promedio de $0,34 \pm 0,01$ en la muestra de todos los alelos. En la muestra de los alelos más frecuentes, el máximo valor de $\overline{D'_{ij}(+)}$ disminuyó sensiblemente (0,72 en el par D11S4185 x D11S935). Este descenso en los valores máximos se refleja también en su media ($0,21 \pm 0,01$). Además, comparando la distribución de $\overline{D'_{ij}(+)}$ entre la muestra de todos los alelos y la muestra con los alelos más frecuentes, se observa claramente como en casi todos los pares de loci disminuye ese valor al eliminar de la muestra los alelos menos frecuentes. Así, en pares de loci como el D11S4185 x D11S4146 y D11S1763 x D11S4148, el promedio de la intensidad de los desequilibrios interalélicos positivos significativos (máximo en la muestra de todos los alelos, en

Resultados

color rojo) no sólo disminuye, sino que pasa a no haber pares de alelos en DG interalélico positivo en la muestra con los alelos más frecuentes.

En los resultados anteriores se observa que los alelos que están contribuyendo a los desequilibrios de alta intensidad son los alelos menos frecuentes. En ocasiones las combinaciones con estos alelos en baja frecuencia son las únicas con DG interalélico positivo significativo de un par de loci. Se ha sugerido que la frecuencia de los alelos está relacionada con su antigüedad en la población (WATTERSON y GUESS 1977; WEBER y WONG 1993). Así, estos alelos en baja frecuencia serían más recientes en la población de manera que la recombinación no habría tenido tiempo de actuar rompiendo las asociaciones.

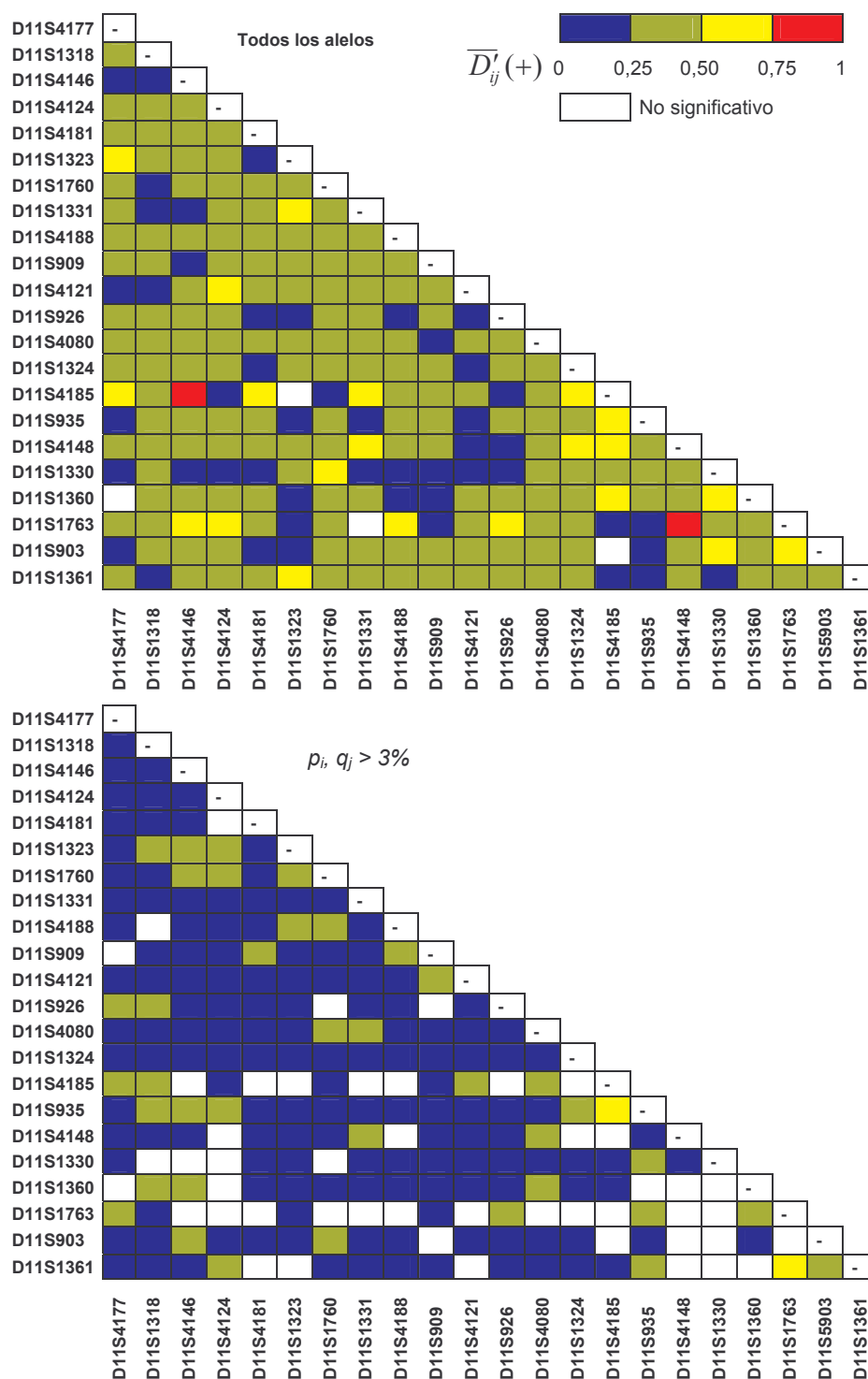


Figura 16. $\overline{D}_{ij}'(+)$ para cada par de microsatélites.

2.2.2. Relación del DG interalélico con el tamaño y frecuencia de los alelos

Se realizó un estudio de la relación del DG con el tamaño y frecuencia de los alelos, para lo cual es necesario clasificarlos en categorías. La tabla 30 y la figura 13 muestran que los alelos en los extremos superior e inferior de las distribuciones de tamaño presentan, en su mayoría, una frecuencia menor del 3% (las excepciones son el alelo 92 de D11S4121 y el alelo 110 de D11S1324) y están separados entre sí por alelos de frecuencia mayor al 3%. Estos alelos de los extremos de la distribución se designaron alelos de tamaño extremo (E) y los restantes se denominaron alelos de tamaño intermedio (I), de acuerdo con ZAPATA *et al.* (2001b). La clasificación por frecuencia fue la misma que la utilizada a lo largo de todo el trabajo. Los alelos con frecuencia menor o igual al 3% se denominaron raros (R), y los alelos con frecuencia mayor al 3% se denominaron comunes (C). Así, cada uno de los haplotipos de dos loci se clasificó por tamaño como EE (formado por dos alelos extremos), EI (formado por un alelo extremo y uno intermedio) y II (formado por dos alelos intermedios). Así mismo, la clasificación por frecuencia dio lugar a haplotipos RR (formados por dos alelos raros), RC (formados por un alelo raro y uno común) y CC (formados por dos alelos comunes).

La tabla 33 muestra que la cantidad de desequilibrios significativos es bastante heterogénea entre los conjuntos de haplotipos. Así, el porcentaje de desequilibrios significativos fue considerablemente mayor en los haplotipos EE (55%), que en los EI (17%) y en los II (14%). Estas diferencias fueron altamente significativas al ser evaluadas mediante tests χ^2 en tablas de contingencia 2 x 2 ($P < 0,001$). La diferencia entre los haplotipos II y EI es significativa ($P < 0,01$), incluso a pesar de que los haplotipos II muestran un desequilibrio solo ligeramente inferior a los haplotipos EI.

Así mismo, se evaluó el porcentaje de desequilibrios según la clasificación de los haplotipos por frecuencia de los alelos, entre los haplotipos RR, RC y CC.

El porcentaje de DG en los haplotipos RR resultó de 52%, una cantidad claramente mayor que en los haplotipos RC (15%) y que en los haplotipos CC (13%). Estas diferencias resultaron altamente significativas ($P \ll 0,001$) al ser evaluadas mediante tests χ^2 en tablas de contingencia 2 x 2. Además, a pesar de la escasa diferencia existente en el porcentaje de desequilibrio entre los haplotipos RC y CC, la probabilidad resultante del test χ^2 resultó sugestiva de ser significativa con un valor justo en el límite del nivel de significación de 0,05.

Tabla 33. Relación del DG con el tamaño y la frecuencia de los alelos			
	Porcentaje de $D'_{ij}(+)$ significativos	χ^2	
Tamaño			
EE	55% (102/187)	EE vs EI:	161,3 ($P \ll 0,001$)
EI	17% (474/2825)	EE vs II:	221,3 ($P \ll 0,001$)
II	14% (635/4503)	EI vs II:	9,7 ($P \ll 0,01$)
Frecuencia			
RR	52% (221/428)	RR vs RC:	344,2 ($P \ll 0,001$)
RC	15% (541/3667)	RR vs CC:	391,7 ($P \ll 0,001$)
CC	13% (449/3417)	RC vs CC:	3,8 ($P = 0,05$)

Estos resultados parecen indicar que los niveles de desequilibrio observados no sólo dependen de las frecuencias alélicas, sino también del tamaño de los alelos implicados. Además, estas diferencias no se pueden atribuir a diferencias en las frecuencias de recombinación ya que éstas resultaron muy similares en los diferentes grupos de haplotipos. Así, por ejemplo las FRE promedio en los grupos RR, RC y CC resultaron $18,6 \pm 0,9$, $17,7 \pm 0,5$ y $19,0 \pm 0,6$; respectivamente.

2.2.3. Desequilibrios globales positivos

La tabla 34 muestra la intensidad de los desequilibrios globales positivos significativos, medida mediante el coeficiente $D'(+)$, para todos los pares de microsatélites.

Resultados

Tabla 34. Intensidad, $D'(+)$, de los desequilibrios globales positivos significativos. Se indica el valor de probabilidad (P) y el valor de significación (α') establecido tras la corrección de Bonferroni.

Pares de loci	Todos los alelos			$p_i, q_j > 3\%$		
	$D'(+) $	P	$\alpha' $	$D'(+) $	P	$\alpha' $
D11S4177 x D11S1318	0,110	0,00110	0,00122	0,098	0,00109	0,00217
D11S4177 x D11S4181				0,046	0,00126	0,00179
D11S4177 x D11S4188	0,106	0,00051	0,00156	0,097	0,00064	0,00278
D11S4177 x D11S4121	0,085	0,00026	0,00100			
D11S4177 x D11S926	0,065	0,00004	0,00152			
D11S4177 x D11S4080	0,099	0,00081	0,00106			
D11S4177 x D11S4148	0,085	0,00052	0,00152			
D11S1318 x D11S4146	0,112	0,00007	0,00106			
D11S1318 x D11S1323	0,127	0,00025	0,00172			
D11S1318 x D11S1760	0,057	0,00025	0,00109	0,045	0,00041	0,00263
D11S1318 x D11S909	0,083	0,00045	0,00132			
D11S1318 x D11S4080	0,084	0,00063	0,00094	0,074	0,00177	0,00208
D11S1318 x D11S1324	0,064	$< 10^{-6}$	0,00094			
D11S1318 x D11S4185	0,126	$< 10^{-6}$	0,00185			
D11S1318 x D11S935	0,103	0,00002	0,00147			
D11S1318 x D11S1361	0,111	0,00003	0,00139	0,104	0,00143	0,00357
D11S4146 x D11S4124	0,103	0,00001	0,00156	0,089	0,00159	0,00333
D11S4146 x D11S4181	0,097	0,00004	0,00104			
D11S4146 x D11S1760				0,117	0,00131	0,00385
D11S4146 x D11S4188	0,119	0,00002	0,00135			
D11S4146 x D11S4121	0,100	0,00105	0,00109	0,095	0,00084	0,00227
D11S4146 x D11S926	0,115	0,00017	0,00156			
D11S4146 x D11S1324	0,072	0,00031	0,00109			
D11S4146 x D11S1763	0,101	0,00014	0,00172			
D11S4146 x D11S903	0,138	0,00117	0,00167	0,132	0,00117	0,00333
D11S4124 x D11S1760	0,120	0,00007	0,00161	0,110	0,00029	0,00385
D11S4124 x D11S909	0,097	0,00155	0,00156			
D11S4124 x D11S1324				0,127	0,00218	0,00294
D11S4124 x D11S4148	0,065	0,00051	0,00167			
D11S4124 x D11S1330	0,114	0,00124	0,00357			
D11S4124 x D11S1361				0,157	0,00781	0,00833
D11S4181 x D11S1323	0,094	0,00095	0,00192	0,089	0,00075	0,00357
D11S4181 x D11S1760	0,091	$< 10^{-6}$	0,00111	0,083	0,00096	0,00250
D11S4181 x D11S909	0,063	0,00026	0,00128			
D11S4181 x D11S4080	0,086	0,00036	0,00086			
D11S4181 x D11S935				0,095	0,00217	0,00294
D11S4181 x D11S4148	0,068	0,00001	0,00135			
D11S4181 x D11S1330				0,140	0,00396	0,00556
D11S4181 x D11S1360	0,131	0,00022	0,00139			
D11S1323 x D11S1331	0,079	0,00139	0,00167			
D11S1323 x D11S4148	0,066	0,00166	0,00294			
D11S1760 x D11S4188	0,101	0,00051	0,00119	0,093	0,00259	0,00278

Tabla 34. (continuación)

Pares de loci	Todos los alelos			$p_i, q_j > 3\%$		
	$D'(+) $	P	α'	$D'(+) $	P	α'
D11S1760 x D11S909	0,119	0,00089	0,00116			
D11S1760 x D11S4080	0,092	0,00063	0,00094			
D11S1760 x D11S1324	0,077	0,00005	0,00106	0,068	0,00003	0,00227
D11S1760 x D11S4148	0,061	$< 10^{-6}$	0,00125			
D11S1331 x D11S4121	0,085	$< 10^{-6}$	0,00091			
D11S1331 x D11S4080	0,103	0,00002	0,00093			
D11S1331 x D11S1360	0,177	0,00002	0,00139	0,166	0,00002	0,00333
D11S1331 x D11S903	0,068	0,00014	0,00122			
D11S4188 x D11S909	0,138	$< 10^{-6}$	0,00172	0,135	$< 10^{-6}$	0,00385
D11S4188 x D11S4080	0,096	0,00022	0,00114			
D11S4188 x D11S1324	0,115	$< 10^{-6}$	0,00143			
D11S4188 x D11S4148	0,058	$< 10^{-6}$	0,00167			
D11S909 x D11S926	0,063	0,00017	0,00167			
D11S909 x D11S1324	0,081	0,00002	0,00128			
D11S4121 x D11S4185	0,141	0,00104	0,00156			
D11S4121 x D11S903	0,088	$< 10^{-6}$	0,00125	0,078	0,00218	0,00227
D11S4121 x D11S1361	0,077	0,00051	0,00125			
D11S926 x D11S4080	0,085	0,00026	0,00128	0,077	0,00026	0,00263
D11S926 x D11S1763	0,104	0,00044	0,00263			
D11S4080 x D11S1324	0,141	0,00002	0,00098	0,134	0,00002	0,00208
D11S4080 x D11S935	0,072	0,00002	0,00135			
D11S4080 x D11S4148	0,075	0,00094	0,00111	0,067	0,00094	0,00294
D11S4080 x D11S1360	0,124	0,00012	0,00119			
D11S4080 x D11S903	0,112	0,00008	0,00122	0,110	0,00080	0,00200
D11S1324 x D11S935	0,130	0,00119	0,00135	0,126	0,00133	0,00313
D11S1324 x D11S1330	0,218	0,00156	0,00278			
D11S1324 x D11S1360	0,099	$< 10^{-6}$	0,00156			
D11S1324 x D11S1763	0,076	0,00037	0,00179			
D11S1324 x D11S903	0,098	0,00005	0,00152			
D11S1324 x D11S1361	0,165	0,00014	0,00139			
D11S4185 x D11S935	0,536	$< 10^{-6}$	0,00313	0,565	$< 10^{-6}$	0,00714
D11S4185 x D11S1361				0,044	0,00585	0,00714
D11S935 x D11S4148	0,112	0,00118	0,00217	0,096	0,00575	0,00625
D11S935 x D11S1330	0,219	0,00141	0,00385	0,213	0,00142	0,00833
D11S4148 x D11S1763	0,080	0,00005	0,00333			
D11S4148 x D11S903	0,064	0,00001	0,00167			
D11S4148 x D11S1361	0,090	$< 10^{-6}$	0,00167			
D11S1330 x D11S1360	0,210	0,00192	0,00417			
D11S1330 x D11S903	0,199	$< 10^{-6}$	0,00385			
D11S1330 x D11S1361	0,088	0,00004	0,00385			
D11S1360 x D11S1763	0,075	0,00237	0,00250			
D11S1360 x D11S903	0,093	0,00022	0,00192	0,088	0,00022	0,00333
D11S1763 x D11S903	0,050	0,00030	0,00263			
D11S1763 x D11S1361	0,073	0,00006	0,00217	0,064	0,00320	0,00714

Resultados

La tabla 35 muestra el porcentaje de casos significativos, así como el rango, media y varianza de las distribuciones de los $D'(+)$ para todos los alelos, y para los alelos más frecuentes. El porcentaje de casos significativos fue de 36% (84/231) para la muestra de todos los alelos y de 14% (33/231) para la muestra de los alelos más frecuentes. La significación estadística se obtuvo mediante el test chi-cuadrado ajustado para múltiples comparaciones con la corrección de Bonferroni (ver *Material y Métodos*). Esta técnica disminuye enormemente el nivel de significación, pudiendo inflar el error de tipo II (probabilidad de aceptar la hipótesis nula cuando es falsa) (ROTHMAN 1990; ZERBA *et al.* 1991) por lo que ofrece una estima muy conservadora de la frecuencia de asociaciones no al azar en esta región. La intensidad de las asociaciones entre pares de microsatélites osciló entre 0,05 (D11S1763 x D11S903) y 0,54 (D11S4185 x D11S935) para todos los alelos y entre 0,04 (D11S4185 x D11S1361) y 0,57 (D11S4185 x D11S935) para los alelos más frecuentes. A pesar de este rango tan amplio, la mayoría de los casos no superaron el valor de 0,2 (95% para la muestra de todos los alelos y 94% para los alelos más frecuentes). Sólo existe un caso (D11S4185 x D11S935) donde el desequilibrio es de intensidad elevada, tanto para la muestra de todos los alelos (0,54), como para la muestra de los alelos más frecuentes (0,57). Así, la media toma un valor bajo tanto para la muestra de todos los alelos ($0,11 \pm 0,01$), como para los alelos más frecuentes ($0,12 \pm 0,02$).

Tabla 35. Estadísticos descriptivos de las distribuciones de los $D'(+)$ significativos. N indica el número de pares de loci en desequilibrio.

	Todos los alelos	Alelos de frecuencia mayor al 3%
N	84	33
%	36% (84/231)	14% (33/231)
Rango	0,05 – 0,54	0,04 – 0,57
Media $\pm$ SE	$0,11 \pm 0,01$	$0,12 \pm 0,02$
Varianza	0,004	0,008

La figura 17 muestra la distribución de frecuencias de la intensidad de los DG globales positivos significativos para todos los alelos y para los alelos más frecuentes. Se observa que la mayoría de las combinaciones corresponden a valores de baja intensidad de desequilibrio. El desequilibrio global es una medida ponderada por la frecuencia haplotípica esperada, de manera que valores altos de desequilibrio interalélico contribuirán en baja medida al valor de desequilibrio global si los alelos implicados presentan frecuencias bajas. Por esta razón los desequilibrios globales positivos significativos toman valores bastante homogéneos y de intensidad baja en su gran mayoría. Así mismo, las bajas intensidades de desequilibrio detectadas demuestran la elevada potencia del test empleado para obtener la significación.

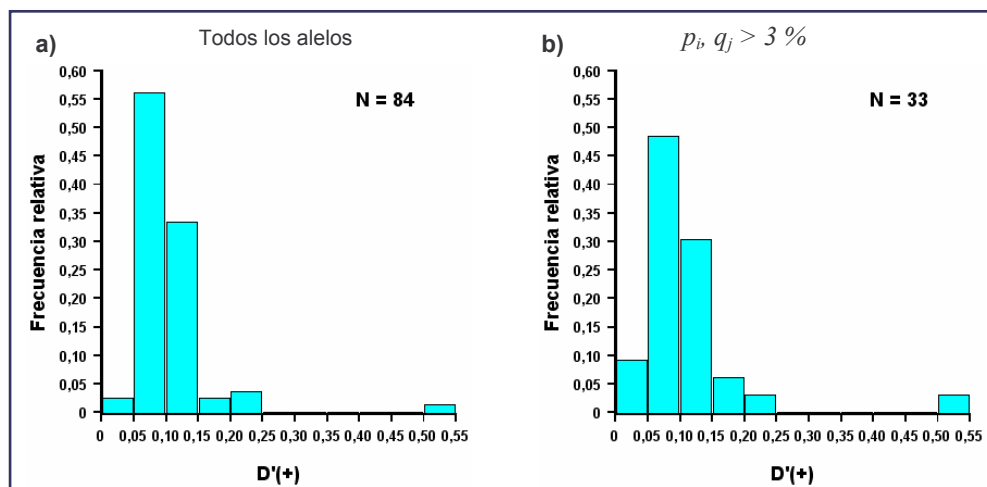


Figura 17. Distribución de las frecuencias de la intensidad [$D'(+)$] de los desequilibrios gaméticos interloci significativos. **a)** Todos los alelos, **b)** alelos más frecuentes ($p_i, q_j > 3\%$). N es el número de pares de microsatélites en desequilibrio significativo.

Si comparamos los niveles obtenidos de desequilibrio global positivo con los desequilibrios interalélicos positivos, se observa una clara disminución de la intensidad, tanto en el rango como en la media (véase tabla 32 para interalélicos). Esta disminución es así mismo debida al efecto de ponderación de los valores de desequilibrio interalélico por sus frecuencias haplotípicas esperadas.

Resultados

La figura 18 muestra a modo de resumen la intensidad del DG global positivo para cada par de microsátélites. Se observa que todos los microsátélites participan en asociaciones no aleatorias, al menos, en alguno de los pares analizados. Así mismo, los desequilibrios están distribuidos a lo largo de toda la región analizada y ningún microsátélite contribuye de forma preferente al desequilibrio detectado, en términos de $D'(+)$.

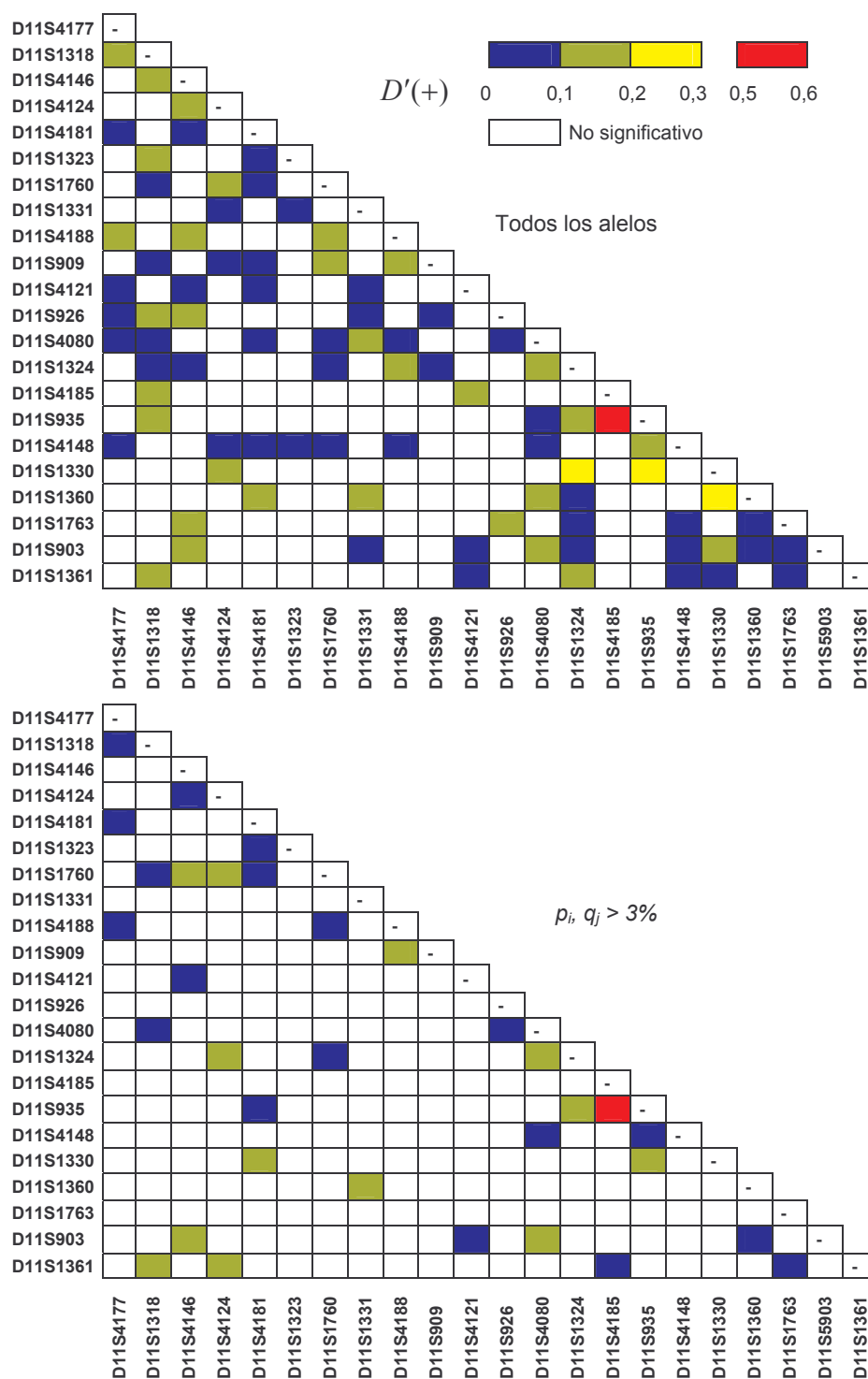


Figura 18. Intensidad de los disequilibrios interloci [$D'(+)$] significativos.

2.2.4. Relación del DG con la FRE y la DF

2.2.4.1. Relación del DG interalélico con la FRE y la DF

Se realizó un análisis de los patrones del DG en función de la FRE y DF. Los 22 microsatélites estudiados se extienden a lo largo de 42,4 cM o 45,5 Mb, con una media de FRE de $20,1 \pm 0,9$ cM y de DF de $19,8 \pm 0,9$ Mb. La tabla 36 muestra el rango y la media de la FRE y de la DF para todos los pares de alelos y para los alelos más frecuentes en desequilibrio significativo. El rango y la media son muy similares en ambos grupos de haplotipos. Se puede observar que el desequilibrio no sólo está confinado a los pares más cercanos (0 cM o 0 Mb), sino que también está presente entre los pares de alelos más alejados (42,4 cM o 45,5 Mb), tanto para la muestra de todos los alelos, como para los alelos más frecuentes. Así, a la vista de estos resultados no podemos establecer un límite máximo de FRE o DF para la existencia de desequilibrio ya que encontramos desequilibrio entre los pares de alelos más alejados de nuestro estudio.

Tabla 36. Rango y media ($\pm$ SE) de la FRE (cM) y DF (Mb) entre los pares de alelos en DG significativo

	Todos los alelos		$p_i, q_j > 3\%$	
	cM	Mb	cM	Mb
Mínimo	0	0	0	0
Máximo	42,4	45,5	42,4	45,5
Media $\pm$ SE	$18,3 \pm 0,4$	$18,0 \pm 0,4$	$19,0 \pm 0,6$	$18,8 \pm 0,7$

En la figura 19 se encuentran representadas las distribuciones de las frecuencias del desequilibrio interalélico positivo significativo en función de la FRE y la DF. Se observa que las distribuciones no parecen seguir un patrón determinado en función de la FRE, aunque en función de la DF parece haber una cierta tendencia hacia un menor número de desequilibrios a medida que ésta aumenta.

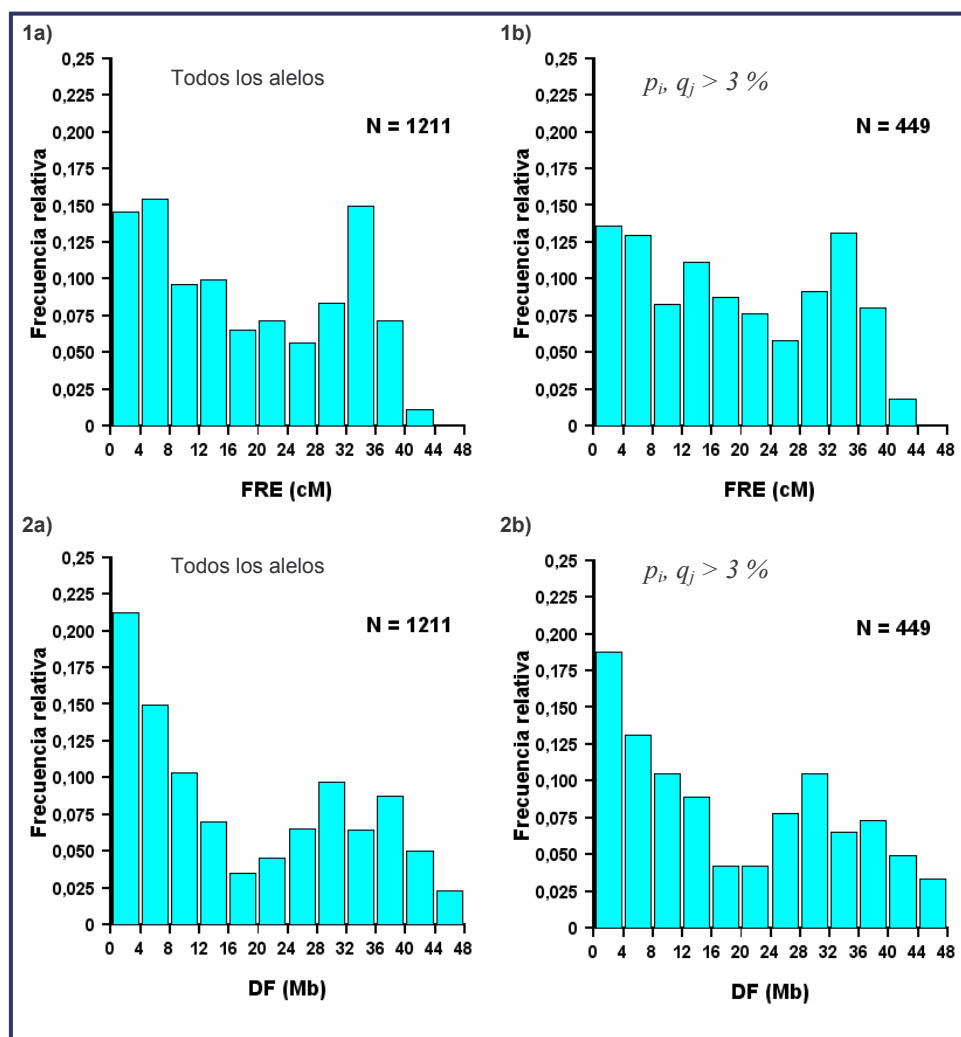


Figura 19. Distribución de las frecuencias de los disequilibrios interalélicos positivos significativos en función de 1) la FRE y 2) la DF, a) para todos los alelos y b) para los alelos más frecuentes ($p_i, q_j > 3\%$).

Para averiguar si existe una cantidad de disequilibrios significativamente mayor entre alelos de microsatélites próximos respecto a los más alejados, se realizó un test χ^2 en tablas de contingencia 2 x 2. Para ello, se dividió el número de disequilibrios interalélicos significativos y no significativos en función de si la FRE es ≤ 2 cM (estrechamente ligados), o si la FRE es > 2 cM (débilmente ligados).

Resultados

Así mismo se realizó este análisis en función de la DF separando el número de desequilibrios significativos y no significativos en función de si la DF es ≤ 2 Mb o si la DF es > 2 Mb. El criterio empleado para separar loci próximos (≤ 2 cM o 2 Mb) de loci alejados (> 2 cM o 2 Mb) ha sido empleado en varias ocasiones para abordar esta misma problemática (PETERSON *et al.* 1995; LAAN y PÄÄBO 1997).

Las tablas 37 y 38 muestran las tablas de contingencia (2 x 2) resultado de agrupar el número de asociaciones significativas y no significativas entre loci próximos (≤ 2 cM o 2 Mb) y entre loci alejados (> 2 cM o 2 Mb). Se observa que la comparación resultó significativa en el caso de todos los alelos, tanto según la FRE ($P < 0,05$) como según la DF ($P < 0,001$). En cambio, al considerar sólo los alelos más frecuentes, la comparación deja de ser significativa tanto para la FRE, como para la DF ($P \gg 0,05$). Estos resultados muestran un comportamiento diferencial de los desequilibrios interalélicos entre los alelos más frecuentes y menos frecuentes con la FRE y DF.

Tabla 37. Test χ^2 junto a su probabilidad (P), de la comparación del número de asociaciones interalélicas significativas y no significativas entre loci estrecha y débilmente ligados, según la FRE.				
	Todos los alelos		$p_i, q_j > 3\%$	
	≤ 2 cM	> 2 cM	≤ 2 cM	> 2 cM
Significativos	92	1119	31	418
No significativos	373	5928	172	2796
	$\chi^2 = 4,9; P = 0,027$		$\chi^2 = 0,9; P = 0,354$	

Tabla 38. Test χ^2 junto a su probabilidad (P), de la comparación del número de asociaciones interalélicas significativas y no significativas entre loci estrecha y débilmente ligados, según la DF.				
	Todos los alelos		$p_i, q_j > 3\%$	
	$\leq 2 \text{ Mb}$	$> 2 \text{ Mb}$	$\leq 2 \text{ Mb}$	$> 2 \text{ Mb}$
Significativos	183	1028	62	387
No significativos	728	5573	348	2620
	$\chi^2 = 12,1; P = 0,001$		$\chi^2 = 1,6; P = 0,205$	

Si los agentes evolutivos actuaran uniformemente a lo largo de la región cromosómica estudiada, cabría esperar que la magnitud del DG decayera a medida que aumenta la FRE y DF. La figura 20 muestra el patrón de estas posibles tendencias, para todos los alelos, y para los alelos más frecuentes ($p_i, q_j > 3\%$). Se observa una gran cantidad de valores de alta intensidad entre loci separados por grandes distancias en la muestra de todos los alelos. Estos desequilibrios corresponden a los alelos en baja frecuencia puesto que en la muestra de alelos más frecuentes están presentes en mucha menor medida.

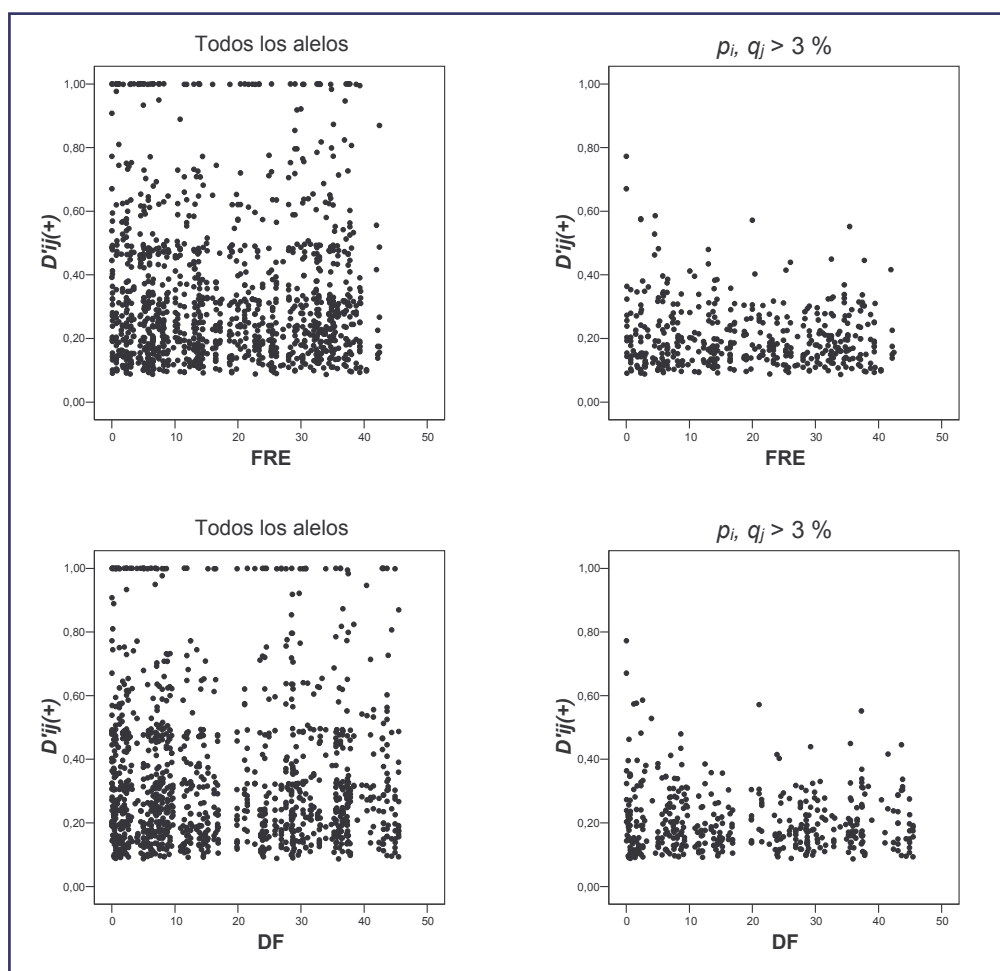


Figura 20. Correlaciones entre la intensidad de los desequilibrios interalélicos positivos, con la FRE y la DF; para todos los alelos, y para los alelos más frecuentes.

Las tablas 39 y 40 muestran el análisis de correlación de los DG interalélicos con la FRE y DF, respectivamente, medida mediante el coeficiente de correlación paramétrico r de Pearson y el coeficiente no paramétrico τ de Kendall.

Tabla 39. Correlación de la intensidad de los desequilibrios interalélicos [$D'_{ij}(+)$] con la **FRE**. Se indica el valor del coeficiente de correlación r de Pearson y τ de Kendall, junto a la probabilidad (P) obtenida mediante el test de Mantel. N indica el número de haplotipos.

	N	r	τ	P
Todos los alelos	1211	-0,037	-0,019	0,096
$p_i, q_j > 3 \%$	449	-0,085	-0,014	0,044

Tabla 40. Correlación de la intensidad de los desequilibrios interalélicos [$D'_{ij}(+)$] con la **DF**. Se indica el valor del coeficiente de correlación r de Pearson y τ de Kendall, junto a la probabilidad (P) obtenida mediante el test de Mantel. N indica el número de haplotipos.

	N	r	τ	P
Todos los alelos	1211	-0,030	-0,019	0,132
$p_i, q_j > 3 \%$	449	-0,083	-0,019	0,216

Se observa una correlación negativa en todos los casos, aunque esta correlación es de escasa magnitud y sólo es significativa ($P < 0,05$) entre el DG y la FRE cuando se eliminan del análisis los alelos en baja frecuencia ($p_i, q_j \leq 3\%$). Este resultado parece indicar que los alelos en baja frecuencia contribuyen a generar desequilibrio independientemente de la distancia que los separa en el cromosoma. Debido a la alta tasa mutacional que poseen los microsatélites, estos alelos podrían tener un origen reciente en la población. Así, la recombinación no habría tenido tiempo de actuar para romper esas asociaciones. Por esta razón, al eliminar los alelos raros, los desequilibrios estarían causados por alelos de cierta antigüedad en la población, de manera que la recombinación actuaría más o menos en función de una mayor o menor FRE o DF entre los alelos en el cromosoma.

Anteriormente se señaló el hecho de que existía una proporción significativamente mayor de desequilibrios interalélicos positivos significativos

Resultados

entre alelos separados por una distancia menor o igual que 2 cM o 2 Mb que separados por más de 2 cM o 2 Mb. Además, la figura 15 parece indicar que existe una tendencia a una disminución en los valores de $D'_{ij}(+)$ al aumentar la FRE y DF a distancias inferiores a 10 cM y 10 Mb. Por esta razón, se realizó un análisis de correlación entre la intensidad de los desequilibrios interalélicos y la FRE y DF para los loci próximos (≤ 2 cM y 2 Mb) y para los loci alejados (> 2 cM y 2 Mb), independientemente.

Las figuras 21 y 22 muestran las correlaciones entre la intensidad del desequilibrio interalélico [$D'_{ij}(+)$] y la FRE y DF, para todos los alelos y para los alelos más frecuentes, respectivamente. Las correlaciones se realizaron para loci estrecha (≤ 2 cM y 2 Mb) y débilmente ligados (> 2 cM y 2 Mb), separadamente. En la figura 21 no se observa ninguna tendencia hacia una disminución del valor de $D'_{ij}(+)$ con una mayor FRE o DF. En cambio en la figura 22 parece existir una ligera tendencia hacia una disminución en la intensidad del desequilibrio interalélico al aumentar la FRE y DF cuando se consideran sólo los pares de microsatélites separados por ≤ 2 cM o 2 Mb en la muestra de los alelos más frecuentes.

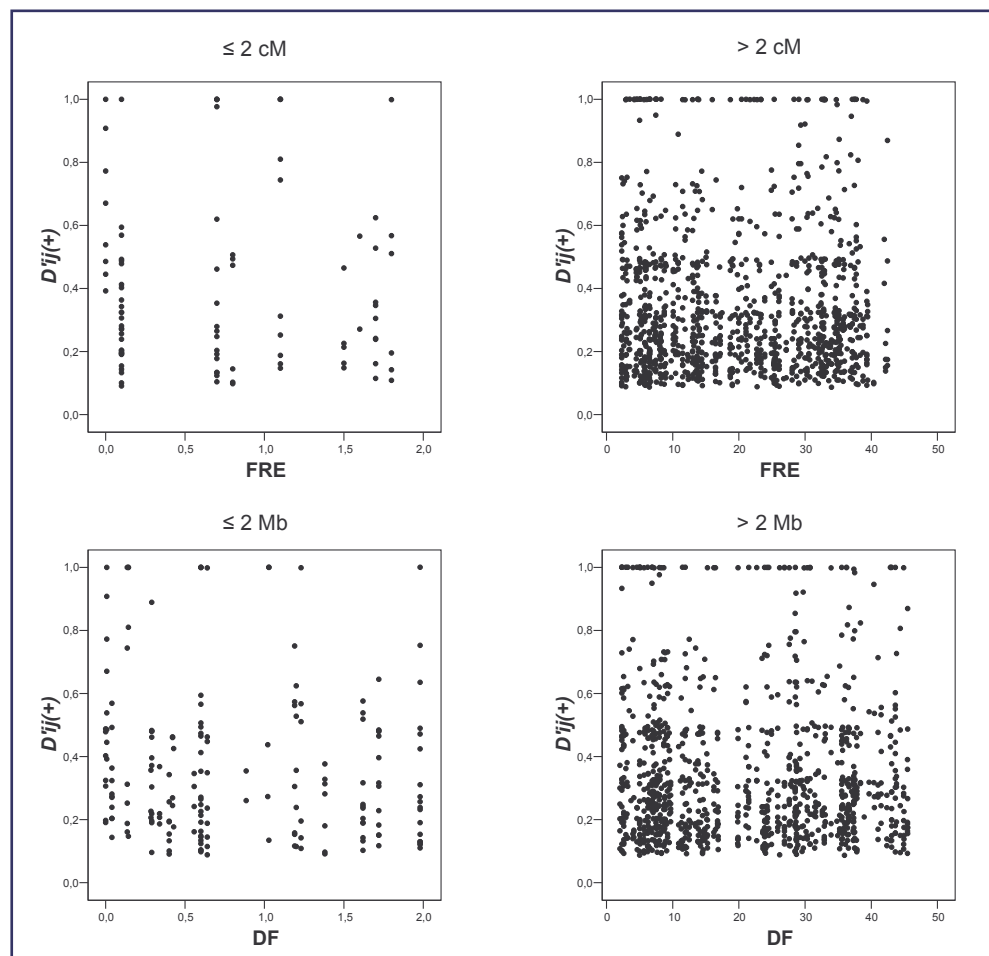


Figura 21. Correlaciones de la intensidad de los desequilibrios interalélicos con la FRE y la DF, para todos los alelos. La correlación se realizó separadamente para pares de loci próximos (≤ 2 cM o 2 Mb) y para pares de loci alejados (> 2 cM o 2 Mb).

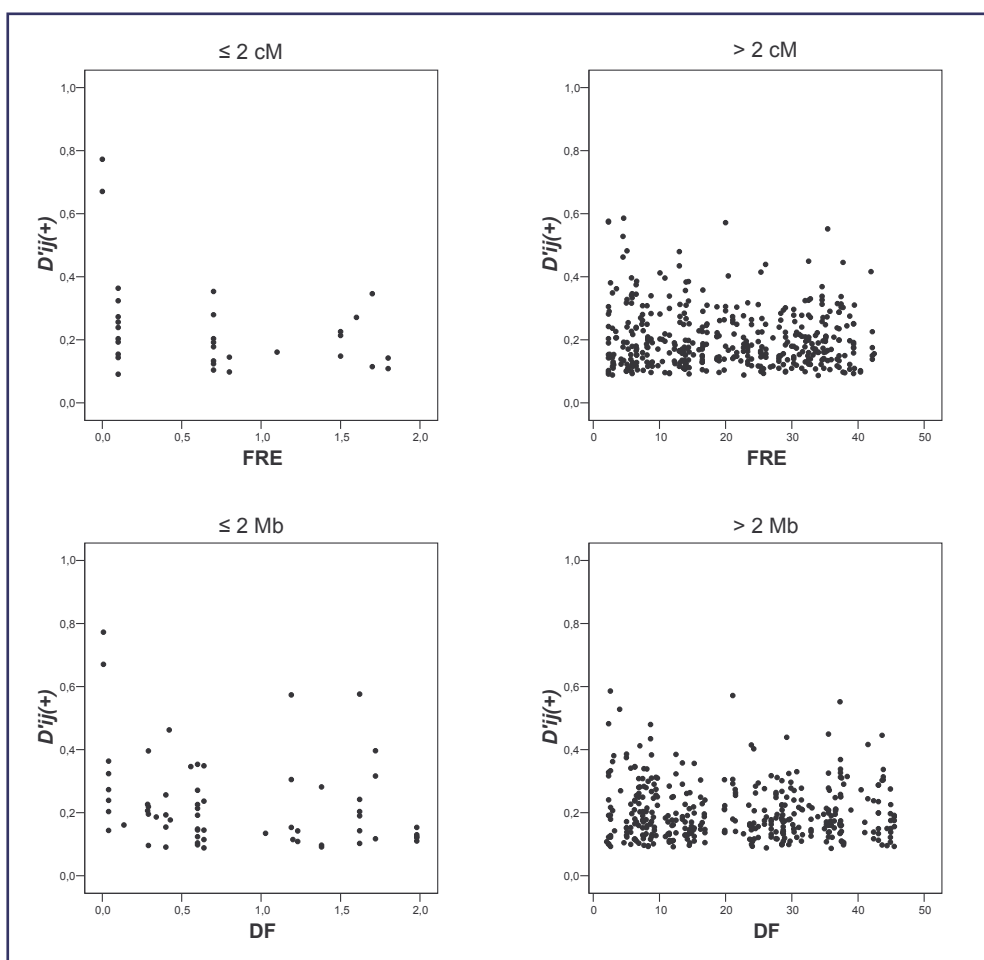


Figura 22. Correlaciones de la intensidad de los disequilibrios interalélicos con la FRE y la DF, para los alelos más frecuentes ($p_i, q_j > 3\%$). La correlación se realizó separadamente para pares de loci próximos (≤ 2 cM o 2 Mb) y para pares de loci alejados (> 2 cM o 2 Mb).

La tabla 41 muestra los resultados del análisis de correlación de la intensidad del disequilibrio interalélico con la FRE para los alelos separados por menos de 2 cM y por más de 2 cM (para todos los alelos y para los alelos más frecuentes). Se observa que las correlaciones muestran signo negativo aunque no son significativas en ningún caso. Si bien los alelos con una $FRE \leq 2$ cM y con una frecuencia mayor del 3%, muestran un valor de probabilidad cercano a la significación ($P = 0,056$). Así mismo, la tabla 42 muestra las correlaciones de la

intensidad del desequilibrio con la DF para los alelos con una DF ≤ 2 Mb y > 2 Mb (todos los alelos y alelos más frecuentes). En este caso la correlación es significativa para los alelos con una DF ≤ 2 Mb y con una frecuencia mayor al 3% ($P = 0,047$). Estos resultados parecen indicar una correlación de la intensidad del desequilibrio interalélico con la distancia entre alelos próximos, cuando se eliminan de la muestra los alelos en baja frecuencia.

Tabla 41. Correlación de la intensidad de los desequilibrios interalélicos [$D'_{ij}(+)$] con la **FRE**, para loci estrecha (≤ 2 cM) y débilmente ligados (> 2 cM). Se indica el valor del coeficiente de correlación r de Pearson y τ de Kendall, junto a la probabilidad (P) obtenida mediante el test de Mantel. N indica el número de haplotipos.

	Todos los alelos				$p_i, q_j > 3 \%$			
	N	r	τ	P	N	r	τ	P
≤ 2 cM	92	-0,033	-0,078	0,379	31	-0,339	-0,277	0,056
> 2 cM	1119	-0,010	-0,005	0,383	418	-0,063	-0,005	0,082

Tabla 42. Correlación de la intensidad de los desequilibrios interalélicos [$D'_{ij}(+)$] con la **DF**, para loci estrecha (≤ 2 Mb) y débilmente ligados (> 2 Mb). Se indica el valor del coeficiente de correlación r de Pearson y τ de Kendall, junto a la probabilidad (P) obtenida mediante el test de Mantel. N indica el número de haplotipos.

	Todos los alelos				$p_i, q_j > 3 \%$			
	N	r	τ	P	N	r	τ	P
≤ 2 Mb	183	-0,116	-0,098	0,070	62	-0,229	-0,229	0,047
> 2 Mb	1028	-0,015	-0,007	0,306	387	-0,049	-0,014	0,194

2.2.4.2. Relación del DG global con la FRE y la DF

Se estudió el comportamiento del desequilibrio global en relación a la FRE y DF. La tabla 43 muestra el mínimo, máximo y media de la FRE y DF entre los pares de microsatélites con asociaciones no al azar significativas. Estos valores sugieren que los desequilibrios están presentes a lo largo de toda la región analizada tanto en la muestra total, como en la muestra de los alelos más frecuentes. Así, el rango de FRE oscila entre 0 y 39 cM para ambas muestras, con una media de $16,1 \pm 1,3$ cM para la muestra de todos los alelos y de $14,3 \pm 2,3$ cM para la muestra de los alelos más frecuentes. Igualmente la DF oscila entre 0 y 45 Mb para ambas muestras, con una media de $15,5 \pm 1,4$ Mb para la muestra de todos los alelos y de $14 \pm 2,6$ Mb para la muestra de los alelos más frecuentes.

Tabla 43. Mínimo, máximo y media de la FRE (cM) y DF (Mb) entre los pares de microsatélites en DG global significativo				
	Todos los alelos		$p_i, q_j > 3\%$	
	cM	Mb	cM	Mb
Mínimo	0	0	0	0
Máximo	39,5	44,9	39,3	44,9
Media $\pm$ SE	$16,1 \pm 1,3$	$15,5 \pm 1,4$	$14,3 \pm 2,27$	$14,0 \pm 2,6$

La figura 23 muestra la distribución de frecuencias de los DG globales significativos en función de la FRE y DF para la muestra de todos los alelos y para la muestra de los alelos más frecuentes. No parece existir una tendencia clara hacia un menor número de asociaciones no al azar entre los microsatélites al aumentar la FRE, aunque existe una cierta tendencia a un menor desequilibrio al aumentar la DF.

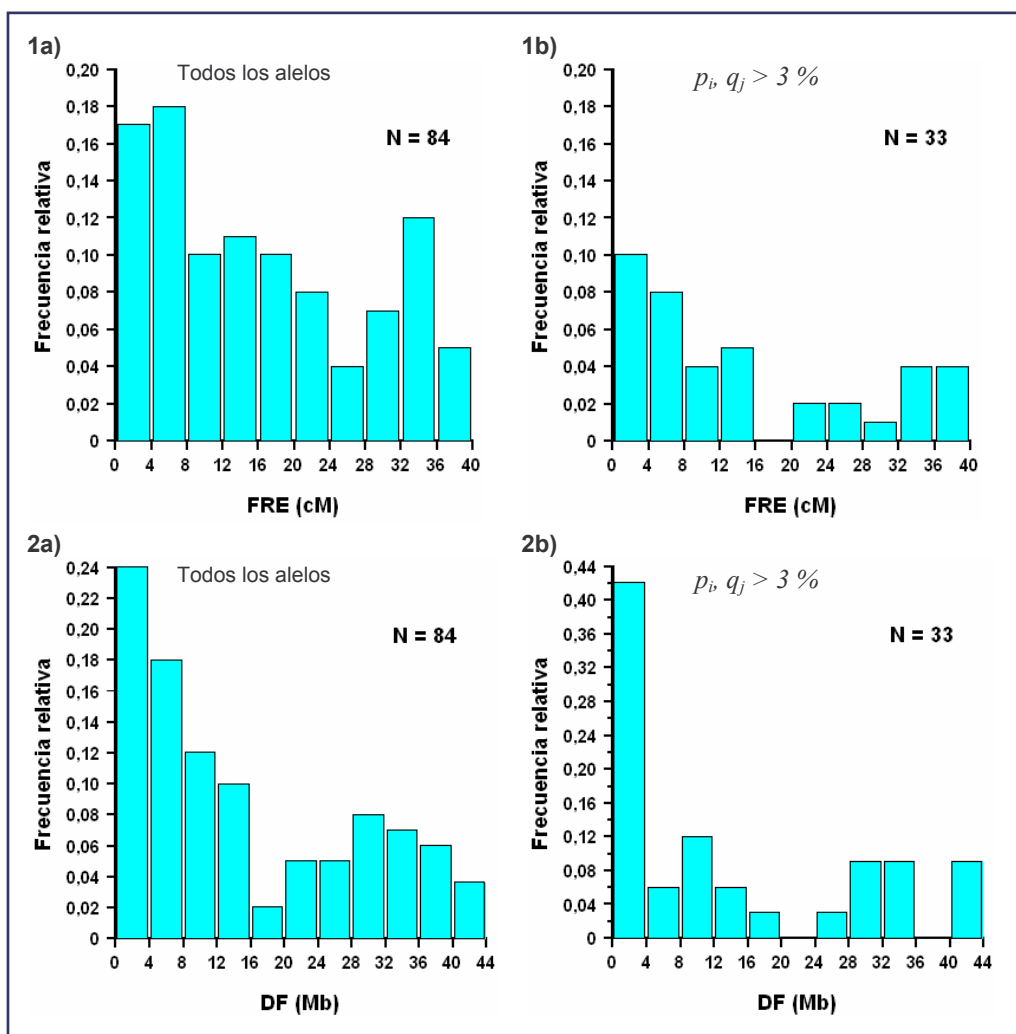


Figura 23. Distribución de las frecuencias de las intensidades de los disequilibrios globales positivos significativos en función de 1) la FRE y 2) la DF, a) para todos los alelos y b) para los alelos más frecuentes ($p_i, q_j > 3\%$).

Para averiguar si la cantidad de disequilibrios globales disminuía con una mayor FRE o DF, se realizó una comparación clasificando el número de asociaciones significativas entre pares de microsatélites, de acuerdo con el mismo

Resultados

criterio arbitrario utilizado para los desequilibrios interalélicos y en estudios previos (PETERSON *et al.* 1995; LAAN y PÄÄBO 1997). Las tablas 44 y 45 muestran los tests χ^2 para las tablas de contingencia 2 x 2 de la comparación del número de asociaciones significativas, frente a las no significativas, entre loci estrechamente ligados (separados por una distancia ≤ 2 cM o 2 Mb) y loci alejados en el cromosoma (separados por una distancia > 2 cM o 2 Mb). La comparación no resultó significativa para la muestra total de alelos, tanto para la FRE, como para la DF. En cambio sí resultó significativa la comparación según la FRE y la DF para la muestra de los alelos más frecuentes. Estos resultados parecen indicar que existe mayor número de pares de microsatélites en desequilibrio significativo entre loci más cercanos que entre los más alejados, al eliminar de la muestra los alelos en baja frecuencia.

Tabla 44. Test χ^2 junto a su probabilidad (P), de la comparación del número de asociaciones globales significativas y no significativas entre loci estrecha y débilmente ligados, según la FRE.

	Todos los alelos		$p_i, q_j > 3\%$	
	≤ 2 cM	> 2 cM	≤ 2 cM	> 2 cM
Significativos	7	77	5	28
No significativos	8	139	10	188
	$\chi^2 = 0,7; P = 0,39$		$\chi^2 = 4,8; P = 0,03$	

Tabla 45. Test χ^2 junto a su probabilidad (P), de la comparación del número de asociaciones globales significativas y no significativas entre loci estrecha y débilmente ligados, según la DF.

	Todos los alelos		$p_i, q_j > 3\%$	
	≤ 2 Mb	> 2 Mb	≤ 2 Mb	> 2 Mb
Significativos	12	72	10	23
No significativos	17	130	19	179
	$\chi^2 = 0,4; P = 0,55$		$\chi^2 = 11,1; P = 0,001$	

Se realizó un análisis de correlación de la intensidad del desequilibrio global $D'(+)$ con la FRE y la DF. En principio cabría esperar una tendencia de decaimiento de la intensidad del desequilibrio con una mayor FRE o DF. La figura 24 muestra las correlaciones de la intensidad del DG global con la FRE y la DF para todos los alelos y para los alelos más frecuentes. Parece existir una tendencia hacia una disminución de la intensidad del DG al aumentar la FRE o DF entre los pares de microsatélites. En cambio, si eliminamos el único valor de alta intensidad de DG que está presente en todos los casos, esta tendencia deja de ser visible.

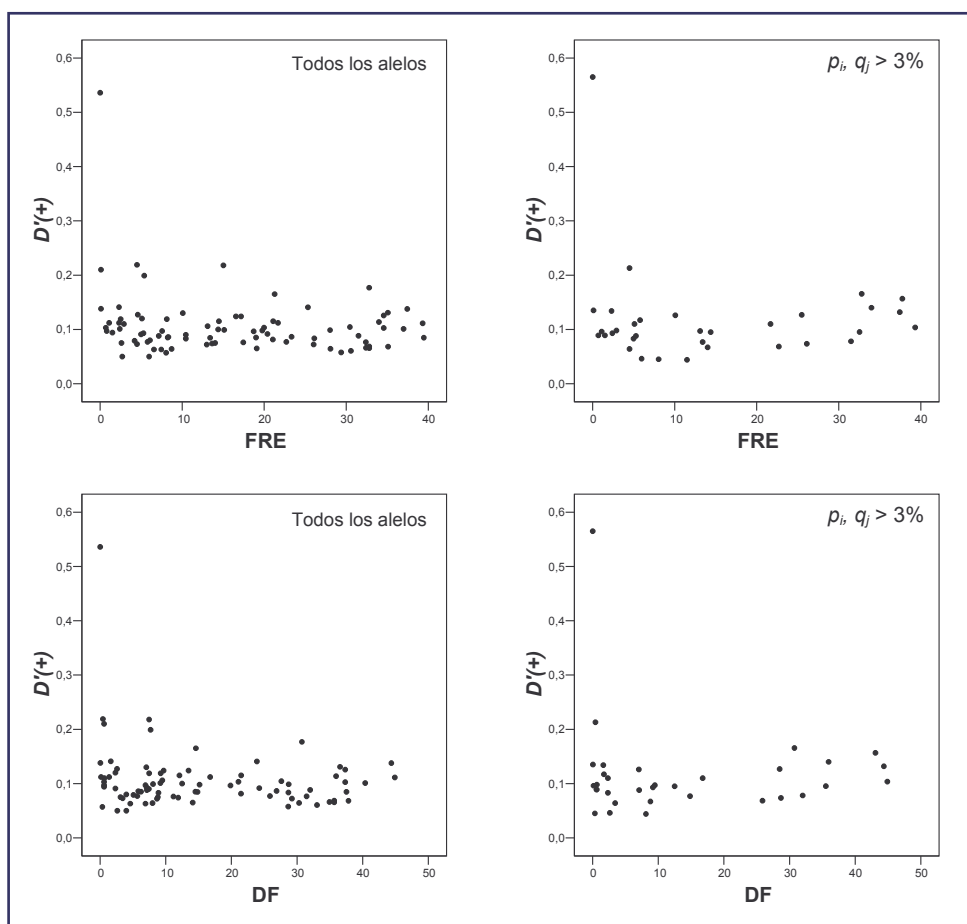


Figura 24. Correlaciones de los desequilibrios globales positivos, con la FRE y la DF; para todos los alelos y para los alelos más frecuentes.

En las tablas 46 y 47 se muestran los coeficientes de correlación. Se observa que toman valores negativos en la mayoría de los casos, siendo significativas las correlaciones con la FRE y DF para la muestra de todos los alelos ($P < 0,05$). Este resultado contrasta con el análisis de los desequilibrios interalélicos, donde sólo se alcanzaba ligeramente la significación en la correlación con la FRE de la muestra de alelos más frecuentes.

Tabla 46. Correlación de la intensidad de los desequilibrios globales con la **FRE**. Se indica el valor del coeficiente de correlación r de Pearson y τ de Kendall, junto a la probabilidad (P) obtenida mediante el test de Mantel. N indica el número de haplotipos.

	N	r	τ	P
Todos los alelos	84	-0,178	-0,083	0,035
$p_i, q_j > 3 \%$	33	-0,095	0,025	0,349

Tabla 47. Correlación de la intensidad de los desequilibrios globales con la **DF**. Se indica el valor del coeficiente de correlación r de Pearson y τ de Kendall, junto a la probabilidad (P) obtenida mediante el test de Mantel. N indica el número de haplotipos.

	N	r	τ	P
Todos los alelos	84	-0,174	-0,072	0,032
$p_i, q_j > 3 \%$	33	-0,061	0,029	0,443

Las figuras 25 y 26 muestran las correlaciones de la intensidad del DG global con la FRE y DF para los loci próximos (≤ 2 cM o 2 Mb), y para loci alejados (> 2 cM o 2 Mb), para todos los alelos y para los alelos más frecuentes, respectivamente. Se observa una tendencia hacia una disminución en la intensidad del DG global al aumentar la FRE y DF entre pares de microsatélites separados por menos de 2 cM o 2 Mb.

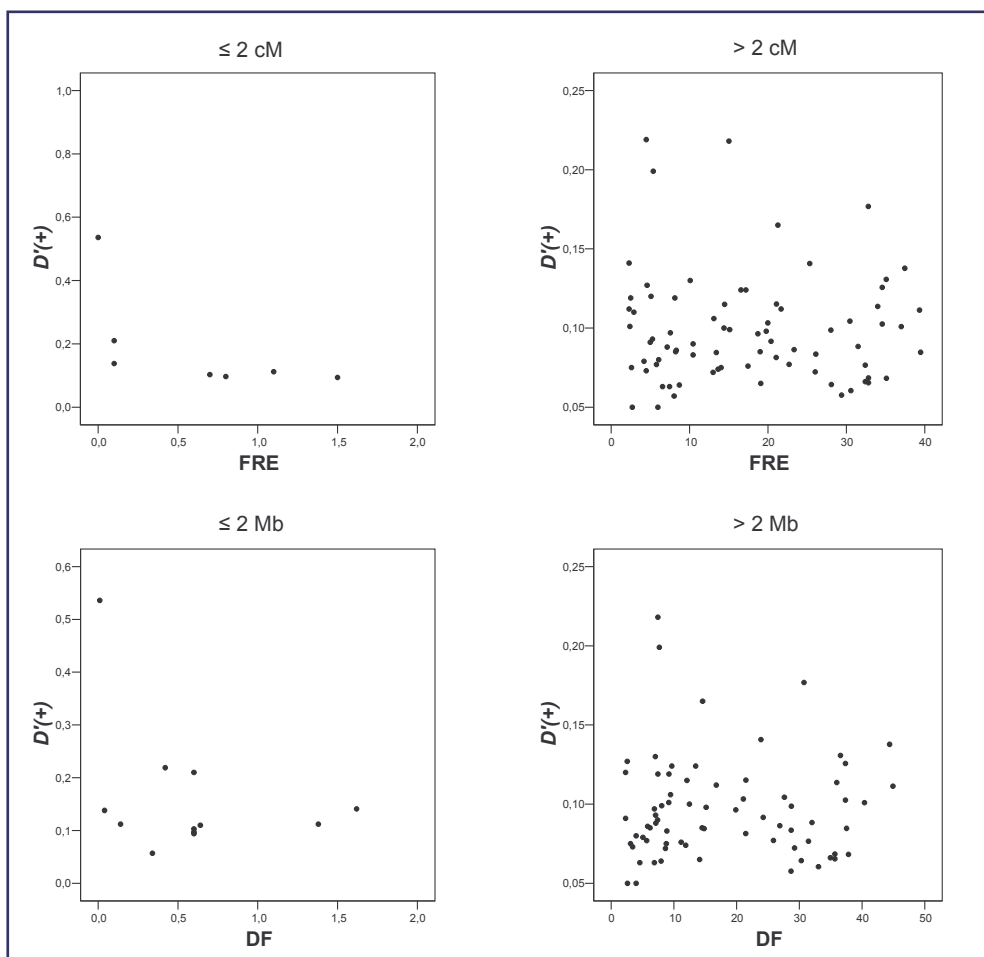


Figura 25. Correlaciones de la intensidad de los DG globales con la FRE y la DF, para todos los alelos. La correlación se realizó separadamente para pares de loci próximos (≤ 2 cM o 2 Mb) y para pares de loci alejados (> 2 cM o 2 Mb).

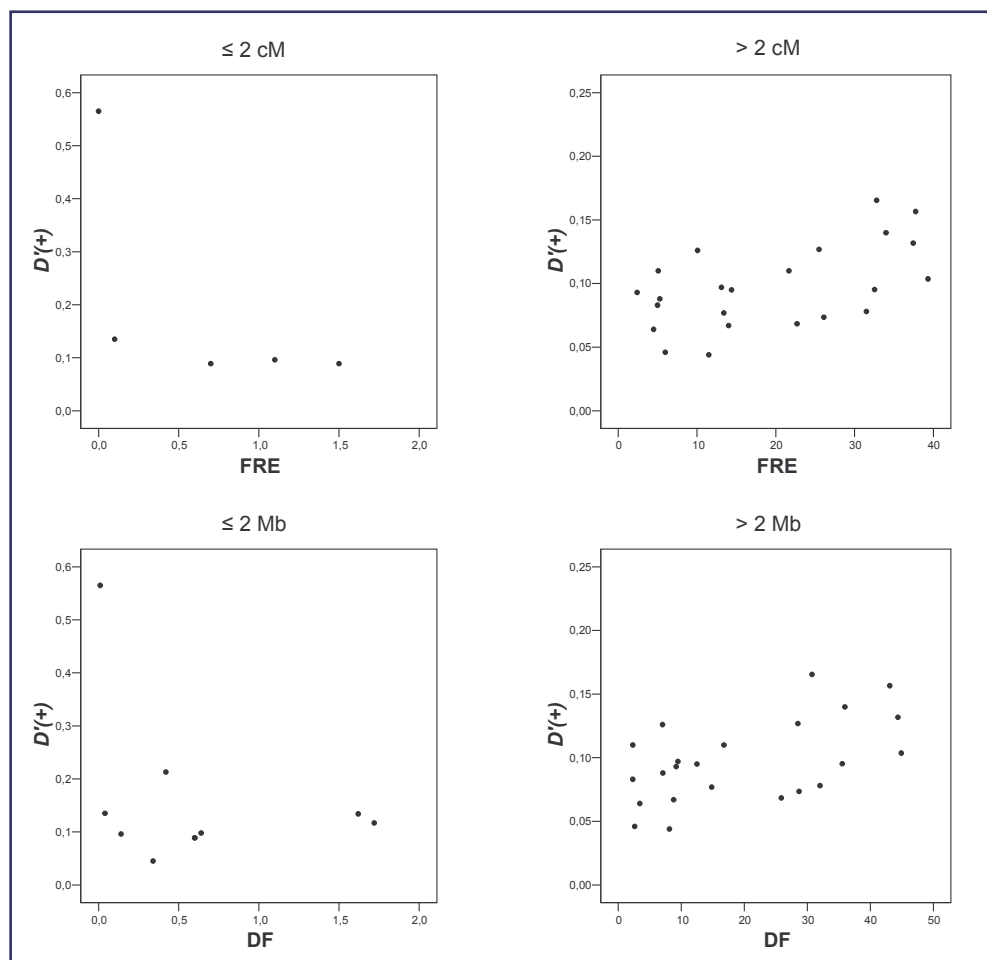


Figura 26. Correlaciones de la intensidad de los DG globales con la FRE y la DF, para los alelos más frecuentes ($p_i, q_j > 3\%$). La correlación se realizó separadamente para pares de loci próximos (≤ 2 cM o 2 Mb) y para pares de loci alejados (> 2 cM o 2 Mb).

Las tablas 48 y 49 muestran el resultado del análisis de correlación entre la intensidad del desequilibrio global y la FRE y DF. Cuando este análisis se realiza sólo entre los pares de microsatélites separados por distancias ≤ 2 cM, la correlación resulta negativa y significativa tanto para la muestra con todos los alelos, como para los alelos más frecuentes. Así mismo, resultó significativa la correlación entre el DG global y la DF entre los pares de microsatélites separados por distancias ≤ 2 Mb, pero en este caso sólo para la muestra que incluye todos

Resultados

los alelos. Las correlaciones entre los pares de microsatélites separados por distancias mayores a 2 cM o 2 Mb no resultaron significativas en ningún caso ($P \gg 0,05$) e, incluso, los coeficientes de correlación presentan signo positivo en la mayoría de los casos. Así, parece que la recombinación ejerce un efecto preponderante disminuyendo la intensidad del DG global al aumentar la FRE entre pares de loci microsatélites. Este efecto sólo se observa cuando se consideran los pares de loci próximos, de manera que a mayores distancias de 2 cM el efecto de la recombinación parece ser superado por otros agentes evolutivos que influyen en el mantenimiento del DG.

Tabla 48. Correlación de la intensidad de los DG globales [$D'(+)$] con la **FRE**, para loci estrecha (≤ 2 cM) y débilmente ligados (> 2 cM). Se indica el valor del coeficiente de correlación r de Pearson y τ de Kendall, junto a la probabilidad (P) obtenida mediante el test de Mantel. N indica el número de haplotipos.

	Todos los alelos				$p_i, q_j > 3 \%$			
	N	r	τ	P	N	r	τ	P
≤ 2 cM	7	-0,624	-0,781	0,016	5	-0,651	-0,738	0,042
> 2 cM	77	-0,027	0,004	0,404	28	0,242	0,149	0,908

Tabla 49. Correlación de la intensidad de los DG globales [$D'(+)$] con la **DF**, para loci estrecha (≤ 2 cM) y débilmente ligados (> 2 cM). Se indica el valor del coeficiente de correlación r de Pearson y τ de Kendall, junto a la probabilidad (P) obtenida mediante el test de Mantel. N indica el número de haplotipos.

	Todos los alelos				$p_i, q_j > 3 \%$			
	N	r	τ	P	N	r	τ	P
≤ 2 Mb	12	-0,337	-0,112	0,049	10	-0,311	-0,136	0,157
> 2 Mb	72	0,015	0,069	0,678	23	0,541	0,369	0,999

II. ESTUDIO DE ASOCIACIÓN DE LA OTOSCLEROSIS CON LOS GENES DEL COLÁGENO TIPO I, COL1A1 Y COL1A2

1. Interpretación de los patrones electroforéticos

En la figura 27 se representan los patrones electroforéticos resultantes de cada uno de los 8 marcadores genéticos analizados. En la parte inferior de las figuras se indica el genotipo de los individuos y en el lateral derecho los tamaños en pares de bases de cada una de las bandas electroforéticas (en el microsatélite (ACT)_n-COL1A2 los números representan el número de repeticiones del trinucleótido). La observación del patrón del microsatélite de repetición dinucleotídico, (GT)_n-COL1A2, constata que nuestro sistema electroforético permite una resolución de hasta 2 pb en microsatélites con repetición dinucleotídica. Además, muestra una elevada precisión en la asignación de tamaños a los alelos de microsatélites y es útil para detectar errores de genotipado que nos llevarían a tipar un heterocigoto como homocigoto debido a una no amplificación total o parcial de uno de los alelos de los microsatélites (RODRÍGUEZ *et al.*, 2001; RODRÍGUEZ & ZAPATA, 2002).

En los SNPs RsaI-COL1A1, EcoRI-COL1A2 y RsaI-COL1A2 aparecen tres bandas en los individuos heterocigotos, dos en los homocigotos +/+ y una en los homocigotos -/-. En estos SNPs la diana de restricción ocupa una posición lateral en el fragmento amplificado. Por esta razón, cuando está presente el alelo (+), la digestión provoca la generación de dos fragmentos de diferente tamaño que se resuelven por separado en el gel de electroforesis. Sin embargo, en el SNP PvuII-COL1A2 la diana de restricción ocupa una posición central en el fragmento amplificado y por tanto el alelo (+) resulta en dos fragmentos de tamaño similar que aparecen como una única banda en la electroforesis. El patrón del SNP MspI-COL1A2 es algo diferente ya que los heterocigotos presentan cuatro bandas electroforéticas. Esto es debido a la presencia de dos dianas de restricción en el fragmento amplificado, de las cuales una es constante y origina dos fragmentos de 600 pb y 1300 pb. La otra diana de restricción es polimórfica y cuando está

Resultados

presente (alelo +), divide el fragmento de 1300 pb en 850 pb + 450 pb. También el fragmento amplificado para el SNP MnlI-COL1A1 tiene más de una diana de restricción, de las cuales una es polimórfica y se resuelve como dos bandas de 195 pb (alelo T) y 134 pb (alelo C).

Como se mencionó en el apartado de *Material y Métodos*, los marcadores (ACT)_n-COL1A2 y EcoRI-COL1A2 permiten conocer la fase gamética en los dobles heterocigotos debido a que se localizan en el mismo fragmento amplificado. En el patrón electroforético de (ACT)/EcoRI-COL1A2 (figura 27), cada dos carriles es el mismo individuo tipado para uno de los dos marcadores genéticos. Esta electroforesis se realizó bajo las mismas condiciones que para el genotipado del microsatélite (ACT)_n-COL1A2. La banda que se observa en el carril del SNP corresponde al alelo (-), ya que las bandas del alelo (+) son de menor tamaño y han migrado a mayor velocidad en la electroforesis de forma que no son visibles en este gel. Se observa que el alelo (-) aparece en todos los casos en fase con el alelo de menor tamaño del microsatélite.

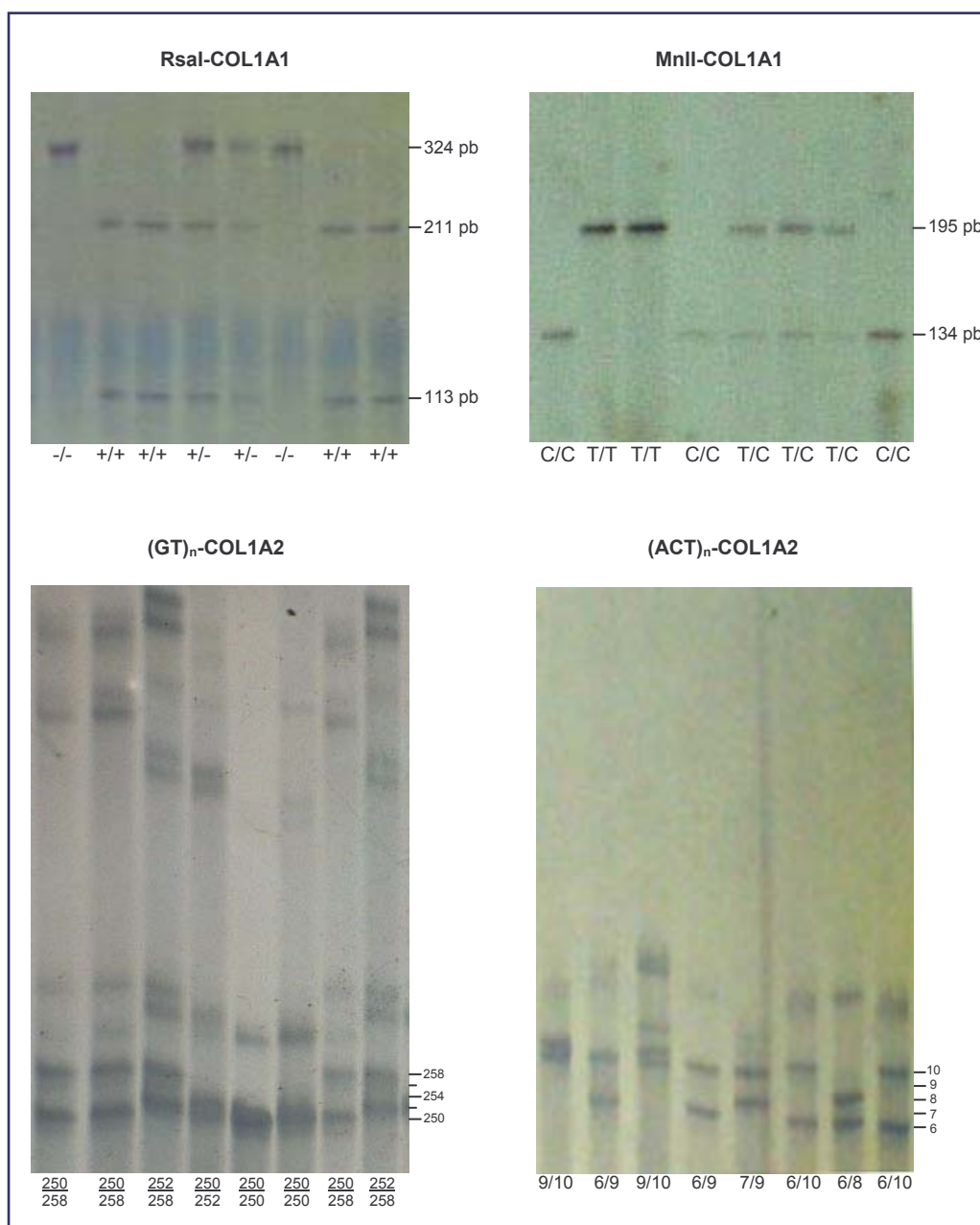


Figura 27. Patrones electroforéticos de los loci analizados. En la parte inferior de cada figura se muestra el genotipo de cada individuo. La numeración en la parte derecha se corresponde con el tamaño de los alelos en pb, salvo en el locus (ACT)_n-COL1A2, donde se representa el número de repeticiones del trinucleótido.

Resultados

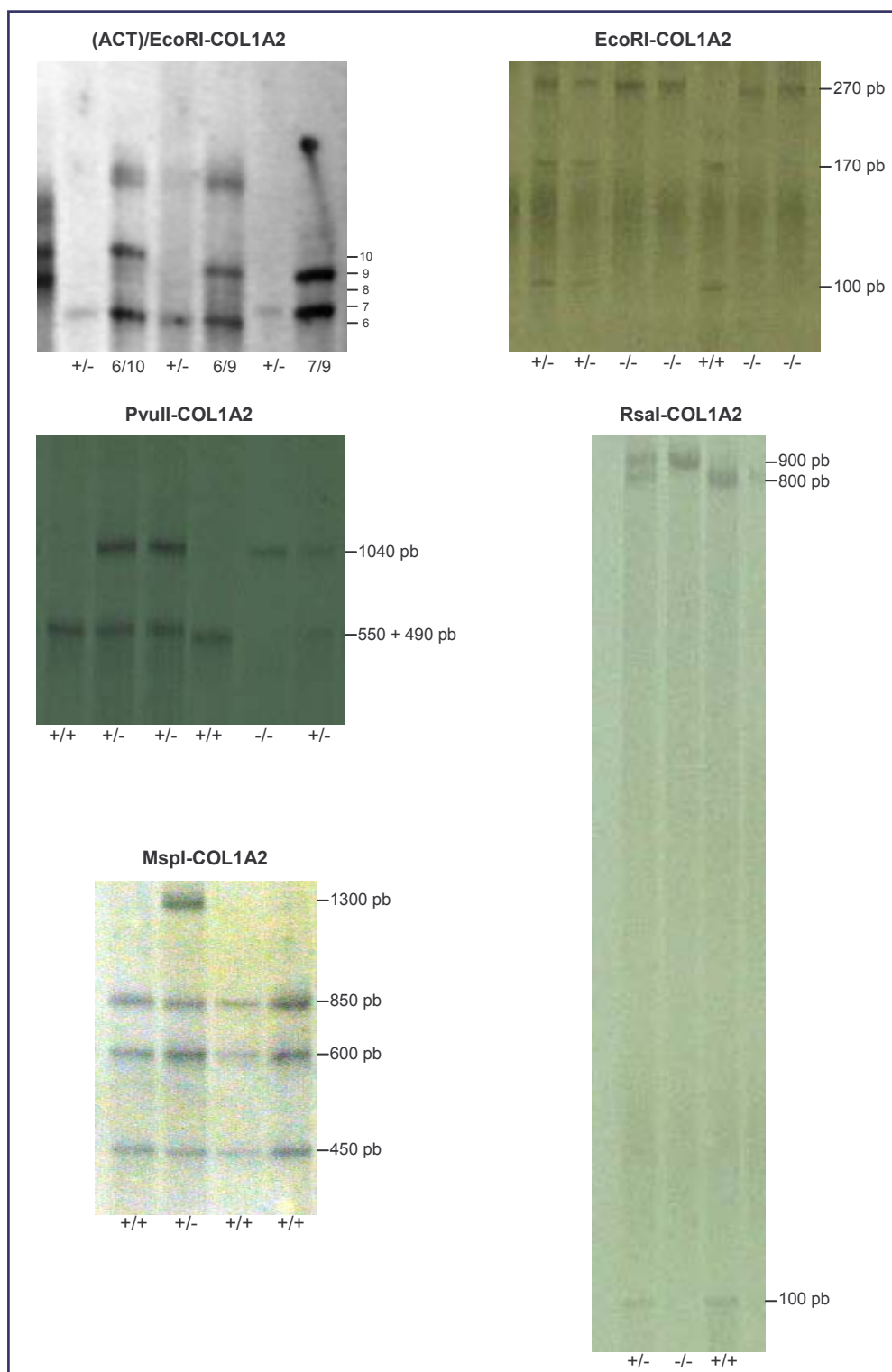


Figura 27. (continuación)

2. Análisis de la variabilidad y estructura genética poblacional

A partir de los datos genotípicos se obtuvieron las frecuencias alélicas para la muestra de casos, controles y la muestra total. En la tabla 50 y la figura 28 se muestran estas frecuencias alélicas y su distribución.

Tabla 50. Frecuencias alélicas de los loci analizados para el estudio de asociación con la otosclerosis, para las muestras control (C), enfermos (D) y la muestra total (C+D)

<i>RsaI</i> /COL1A1			<i>EcoRI</i> /COL1A2			<i>RsaI</i> /COL1A2		
Alelo	Grupo	Frec	Alelo	Grupo	Frec	Alelo	Grupo	Frec
+	C	0,760	+	C	0,350	+	C	0,760
	D	0,750		D	0,295		D	0,720
	C+D	0,755		C+D	0,323		C+D	0,740
-	C	0,240	-	C	0,650	-	C	0,240
	D	0,250		D	0,705		D	0,280
	C+D	0,245		C+D	0,677		C+D	0,260
<i>MnlI</i> /COL1A1			<i>(ACT)_n</i> /COL1A2			<i>MspI</i> /COL1A2		
Alelo	Grupo	Frec	Alelo	Grupo	Frec	Alelo	Grupo	Frec
C	C	0,690	6	C	0,290	+	C	0,920
	D	0,730		D	0,255		D	0,945
	C+D	0,710		C+D	0,273		C+D	0,933
T	C	0,310	7	C	0,195	-	C	0,080
	D	0,270		D	0,220		D	0,055
	C+D	0,290		C+D	0,207		C+D	0,067
<i>(GT)_n</i> /COL1A2			8	C	0,170			
Alelo	Grupo	Frec		D	0,235			
250	C	0,470		C+D	0,203			
	D	0,445	9	C	0,145			
	C+D	0,458		D	0,120			
252	C	0,310		C+D	0,132			
	D	0,285	10	C	0,200			
	C+D	0,297		D	0,170			
254	C	0,005		C+D	0,185			
	D	0,010						
	C+D	0,008						
256	C	0,005	<i>PvuII</i> /COL1A2					
	D	0,005	Alelo	Grupo	Frec			
	C+D	0,005	+	C	0,635			
258	C	0,200		D	0,615			
	D	0,250		C+D	0,625			
	C+D	0,225	-	C	0,365			
260	C	0,010		D	0,385			
	D	0,005		C+D	0,375			
	C+D	0,007						

Resultados

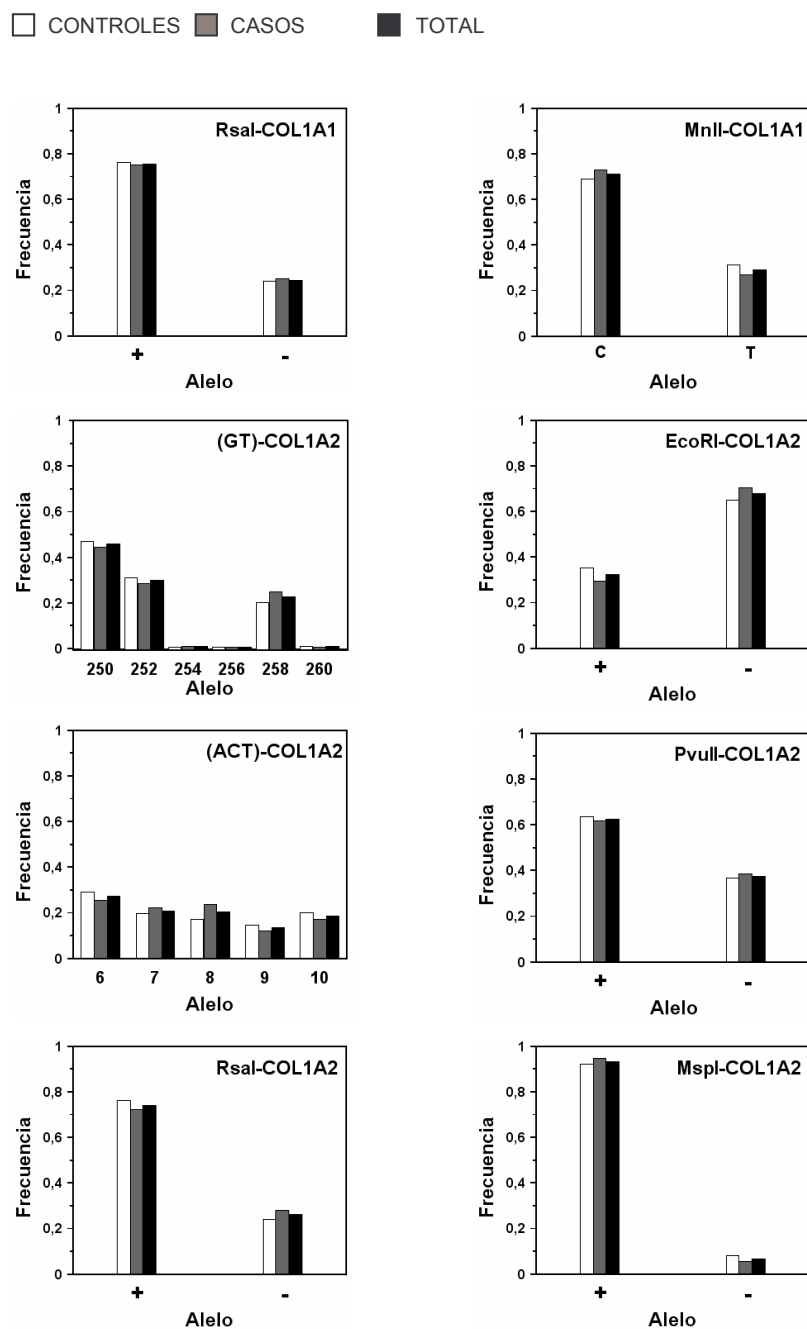


Figura 28. Distribución de las frecuencias alélicas de los marcadores estudiados en los dos grupos de individuos (casos y controles) y en la muestra total.

Se observa que no existen diferencias apreciables en las frecuencias alélicas entre los grupos de casos, controles y la muestra total.

La tabla 51 resume las estimas de variabilidad, en la muestra control, en la muestra de casos y en la muestra total, para cada locus analizado. Así como la magnitud y significación de las desviaciones de las proporciones H-W para cada locus medido mediante el estimador f_T y el test de la chi-cuadrado asociado, respectivamente (ROBERTSON y HILL, 1984).

Resultados

Tabla 51. Magnitud ($\hat{f}$) y significación (χ^2) de las desviaciones de las proporciones H-W. Para las muestras control, C, muestra de enfermos, D y muestra total, C + D (casos + controles).

Marcador/Locus	Grupo	Nº alelos	Heterozigosis observada	Heterozigosis esperada	$\hat{f} \pm SE$	χ^2	P
RsaI / COL1A1	C	2	0,420	0,367	-0,147 $\pm$ 0,084	2,16	0,14
	D	2	0,380	0,377	-0,008 $\pm$ 0,099	0,01	0,93
	C+D	2	0,400	0,370	-0,079 $\pm$ 0,066	1,25	0,26
MnII / COL1A1	C	2	0,460	0,430	-0,071 $\pm$ 0,097	0,50	0,48
	D	2	0,360	0,396	0,092 $\pm$ 0,104	0,85	0,36
	C+D	2	0,410	0,412	0,007 $\pm$ 0,071	0,01	0,92
(GT) _n / COL1A2	C	6	0,620	0,646	0,008 $\pm$ 0,051	0,03	0,86
	D	6	0,630	0,661	0,021 $\pm$ 0,071	0,21	0,64
	C+D	6	0,625	0,651	0,015 $\pm$ 0,046	0,22	0,64
EcoRI / COL1A2	C	2	0,400	0,457	0,127 $\pm$ 0,101	1,60	0,21
	D	2	0,410	0,418	0,019 $\pm$ 0,100	0,04	0,85
	C+D	2	0,405	0,437	0,076 $\pm$ 0,072	1,15	0,28
(ACT) _n / COL1A2	C	5	0,830	0,792	-0,054 $\pm$ 0,044	1,17	0,28
	D	5	0,820	0,790	-0,030 $\pm$ 0,047	0,37	0,54
	C+D	5	0,825	0,790	-0,043 $\pm$ 0,032	1,47	0,23
PvuII / COL1A2	C	2	0,470	0,466	-0,009 $\pm$ 0,100	0,01	0,93
	D	2	0,510	0,476	-0,072 $\pm$ 0,099	0,52	0,47
	C+D	2	0,490	0,469	-0,043 $\pm$ 0,070	0,37	0,54
RsaI / COL1A2	C	2	0,380	0,367	-0,037 $\pm$ 0,097	0,14	0,71
	D	2	0,420	0,405	-0,037 $\pm$ 0,098	0,14	0,71
	C+D	2	0,400	0,385	-0,037 $\pm$ 0,069	0,28	0,60
MspI / COL1A2	C	2	0,160	0,148	-0,082 $\pm$ 0,022	0,68	0,41
	D	2	0,110	0,104	-0,053 $\pm$ 0,018	0,29	0,59
	C+D	2	0,135	0,126	-0,070 $\pm$ 0,014	0,98	0,32
Media $\pm$ S.E.	C	2,9 $\pm$ 0,6	0,468 $\pm$ 0,069	0,459 $\pm$ 0,068	0,067 $\pm$ 0,018		
	D	2,9 $\pm$ 0,6	0,455 $\pm$ 0,074	0,453 $\pm$ 0,064	0,042 $\pm$ 0,010		
	C+D	2,9 $\pm$ 0,6	0,461 $\pm$ 0,071	0,450 $\pm$ 0,072	0,046 $\pm$ 0,010		

Las medias se calcularon a partir de las f en valor absoluto

De la comparación de la heterozigosis observada en las tres muestras (casos, controles y total) resultan pocas diferencias. La heterozigosis observada en la muestra control varía de 0,16 en el locus *MspI/COL1A2* a 0,83 en el locus *(ACT)_n/COL1A2*, con un valor promedio de $0,47 \pm 0,069$. Así mismo, la heterozigosis observada en la muestra de casos varía de 0,11 (*MspI/COL1A2*) a 0,82 [*(ACT)_n/COL1A2*], con un valor promedio de $0,46 \pm 0,074$. Además la heterozigosis observada en el grupo de individuos combinados (muestra total)

también tiene su máximo valor en el locus $(ACT)_n/COL1A2$ (0,83) y mínimo en el locus $MspI/COL1A2$ (0,14), con un promedio de $0,46 \pm 0,071$. El hecho de que el valor más elevado de heterozigosis corresponda al locus $(ACT)_n/COL1A2$ se debe a su carácter multialélico y a que las frecuencias alélicas de los distintos alelos fueron muy similares (figura 28), a diferencia del locus $(GT)_n/COL1A2$ que, teniendo un mayor número de alelos, presenta una menor heterozigosis observada debido a las importantes diferencias en frecuencia de una mitad de sus alelos frente a la otra mitad.

La comparación entre la heterozigosis observada y esperada, para cada locus dentro de cada muestra, refleja en general una buena concordancia para los tres grupos de individuos. En la muestra control, el promedio de la heterozigosis observada resultó ser de $0,47 \pm 0,07$ y el promedio de la heterozigosis esperada, $0,46 \pm 0,07$. En la muestra de casos, el promedio de la heterozigosis observada resultó ser de $0,46 \pm 0,07$ y el promedio de la heterozigosis esperada, $0,45 \pm 0,06$. Respecto a la muestra total, donde tampoco se detectan diferencias, el promedio de la heterozigosis observada fue de $0,46 \pm 0,07$ frente a un promedio de $0,45 \pm 0,07$ en la heterozigosis esperada. En consecuencia, aplicando el test χ^2 asociado al estimador f_T y f_{ii} (ROBERTSON Y HILL, 1984), no se detectan diferencias estadísticamente significativas entre los valores genotípicos observados y los esperados bajo las proporciones H-W dentro de las muestras analizadas ($P > 0,05$) (tabla 47). La falta de potencia estadística, debida a desviaciones pequeñas o moderadas de las proporciones H-W y/o a pequeños tamaños muestrales, es un obstáculo en la detección de subestructuración poblacional. Sin embargo, en este análisis no cabe considerar un posible problema de potencia, dado que la magnitud de las desviaciones de las proporciones H-W medidas mediante estadísticos f , son generalmente pequeñas. Así, en la muestra control oscilan (en valor absoluto) entre $0,009 \pm 0,100$ ($PvuII/COL1A2$) y $0,147 \pm 0,084$ ($RsaI/COL1A1$), con un promedio de $0,067 \pm 0,018$. En la muestra de casos los valores absolutos de las estimas de $\hat{f}_T$ oscilan entre $0,008 \pm 0,099$ ($RsaI/COL1A1$) y $0,092 \pm 0,104$ ($MnII/COL1A1$), con un promedio de $0,047 \pm$

Resultados

0,010. Lo mismo ocurre en la muestra total donde los valores varían entre $0,007 \pm 0,071$ (*MnII/COL1A1*) y $0,079 \pm 0,066$ (*RsaI/COL1A1*) con un valor promedio de $0,046 \pm 0,010$. Así mismo tampoco se ha presentado una tendencia significativa hacia un exceso o un defecto de homocigotos mediante un test del signo ($P > 0,05$).

Para los loci multialélicos se pueden calcular las desviaciones de las proporciones Hardy-Weinberg por alelo. En la tabla 52 se muestran la magnitud y significación de las desviaciones H-W por alelo. Se observa que las desviaciones son de baja magnitud y en ningún caso estadísticamente significativas. Véase que en el grupo control la magnitud de las desviaciones (en valor absoluto), medidas mediante el estimador f_{ii} , varían entre 0,000 (alelos 254 y 256 de *(GT)_n/COL1A2*) y 0,147 (alelos de *RsaI/COL1A1*), con un valor promedio (de los valores absolutos) de $0,062 \pm 0,010$. También en la muestra de enfermos la magnitud de las desviaciones es baja y oscila entre 0,000 (alelos 256 y 260 de *(GT)_n/COL1A2*), y 0,192 (alelo 8 de *(ACT)_n/COL1A2*), con un promedio de $0,055 \pm 0,011$. Respecto a la muestra de individuos agrupados (casos y controles), la magnitud de las desviaciones en términos de $|f_{ii}|$ varía entre 0,003 (alelo 256 del locus *(GT)_n/COL1A2*) y 0,128 (alelo 8 del locus *(ACT)/COL1A2*) con un promedio de $0,045 \pm 0,007$.

De todo lo anteriormente expuesto se puede concluir que las frecuencias genotípicas están de acuerdo en todos los casos con las proporciones esperadas bajo equilibrio H-W en las muestras control, enfermos y en la muestra combinada (casos y controles agrupados). Por lo tanto podemos excluir la existencia de algún fenómeno de subestructuración poblacional en nuestro estudio.

Tabla 52. Magnitud ($\hat{f}_{ii}$) y significación (X^2) de las desviaciones Hardy-Weinberg por alelo para cada uno de los loci analizados en las muestras control (C), enfermos (D) y en la muestra total (C+D).

RsaI / COL1A1					(ACT) _n / COL1A2				
Alelo	Grupo	$\hat{f}_{ii}$	X ²	P	Alelo	Grupo	$\hat{f}_{ii}$	X ²	P
+	C	-0,147	2,16	0,14	6	C	-0,015	0,02	0,88
	D	-0,008	0,01	0,93		D	-0,074	0,55	0,46
	C+D	-0,079	1,25	0,26		C+D	-0,044	0,39	0,53
-	C	-0,147	2,16	0,14	7	C	-0,110	1,22	0,27
	D	-0,008	0,01	0,93		D	0,132	1,73	0,19
	C+D	-0,079	1,25	0,26		C+D	0,014	0,04	0,84
MnII / COL1A1					8	C	-0,058	0,34	0,56
Alelo	Grupo	$\hat{f}_{ii}$	X ²	P		D	-0,192	3,68	0,06
C	C	-0,071	0,50	0,48		C+D	-0,128	3,27	0,07
	D	0,092	0,85	0,36	9	C	-0,084	0,71	0,40
	C+D	0,007	0,01	0,92		D	-0,037	0,14	0,71
T	C	-0,071	0,50	0,48		C+D	-0,063	0,80	0,37
	D	0,092	0,85	0,36	10	C	0,005	0,00	0,96
	C+D	0,007	0,01	0,92		D	0,013	0,02	0,90
C+D						C+D	0,008	0,01	0,91
(GT) _n / COL1A2					PvuII / COL1A2				
Alelo	Grupo	$\hat{f}_{ii}$	X ²	P	Alelo	Grupo	$\hat{f}_{ii}$	X ²	P
250	C	0,082	0,67	0,41	+	C	-0,009	0,01	0,93
	D	0,054	0,29	0,59		D	-0,072	0,52	0,47
	C+D	0,066	0,87	0,35		C+D	-0,043	0,37	0,54
252	C	0,070	0,50	0,48	-	C	-0,009	0,01	0,93
	D	-0,050	0,25	0,62		D	-0,072	0,52	0,47
	C+D	0,010	0,02	0,89		C+D	-0,043	0,37	0,54
254	C	0,000	-	-	RsaI / COL1A2				
	D	-0,005	0,00	0,96	Alelo	Grupo	$\hat{f}_{ii}$	X ²	P
	C+D	-0,005	0,01	0,94	+	C	-0,037	0,14	0,71
256	C	0,000	-	-		D	-0,037	0,14	0,71
	D	0,000	-	-		C+D	-0,037	0,28	0,60
	C+D	-0,003	0,00	0,97	-	C	-0,037	0,14	0,71
258	C	-0,058	0,33	0,56		D	-0,037	0,14	0,71
	D	0,152	2,32	0,13		C+D	-0,037	0,28	0,60
	C+D	0,056	0,64	0,43	MspI / COL1A2				
260	C	-0,005	0,00	0,96	Alelo	Grupo	$\hat{f}_{ii}$	X ²	P
	D	0,000	-	-	+	C	-0,082	0,68	0,41
	C+D	-0,005	0,01	0,94		D	-0,053	0,29	0,59
C+D				C+D		-0,070	0,98	0,32	
EcoRI / COL1A2					-	C	-0,082	0,68	0,41
Alelo	Grupo	$\hat{f}_{ii}$	X ²	P		D	-0,053	0,29	0,59
+	C	0,127	1,60	0,21		C+D	-0,070	0,98	0,32
	D	0,019	0,04	0,85	-	C	-0,082	0,68	0,41
	C+D	0,076	1,15	0,28		D	-0,053	0,29	0,59
-	C	0,127	1,60	0,21		C+D	-0,070	0,98	0,32
	D	0,019	0,04	0,85					
	C+D	0,076	1,15	0,28					

3. Análisis de asociación

En la figura 28 se puede observar que las frecuencias alélicas son similares en casos y controles. Además, la comparación de estas frecuencias mediante tests χ^2 de homogeneidad no muestra diferencias estadísticamente significativas en ningún caso ($P > 0,05$). La intensidad de asociación entre los alelos y la otosclerosis medida en términos de coeficientes D' se muestra en la tabla 53. Los valores oscilaron (en valor absoluto) entre 0,02 (*RsaI* / *COL1A1*) y 0,19 (*MspI* / *COL1A2*), con un valor medio de $0,08 \pm 0,02$. Este promedio está muy cercano al mínimo valor que puede tener D' . Sólo la asociación entre *MspI* y la otosclerosis ($D' = 0,19 \pm 0,18$) muestra una desviación considerable de la asociación aleatoria, probablemente debida a un error de muestreo. Se debe hacer notar que el error de muestreo en D' se hace más pronunciado cuando uno de los marginales de la tabla de contingencia 2 x 2 tiene una frecuencia intermedia, mientras la frecuencia del otro marginal se desvía considerablemente de 0,5 (ZAPATA *et al.*, 1997). Este es precisamente el caso para la tabla de *MspI* y la otosclerosis.

Tabla 53. Asociación de los alelos de cada marcador con la otosclerosis

Marcador / Locus	$D' (\pm SE)$	χ^2	g.l.	P
<i>RsaI</i> / <i>COL1A1</i>	$-0,020 \pm 0,088$	0,05	1	0,82
<i>MnII</i> / <i>COL1A1</i>	$0,070 \pm 0,078$	0,78	1	0,38
(GT) _n / <i>COL1A2</i>	$-0,055 \pm 0,035$	2,12	5	0,83
<i>EcoRI</i> / <i>COL1A2</i>	$-0,085 \pm 0,072$	1,38	1	0,24
(ACT) _n / <i>COL1A2</i>	$-0,090 \pm 0,045$	3,80	4	0,43
<i>PvuII</i> / <i>COL1A2</i>	$-0,027 \pm 0,065$	0,17	1	0,24
<i>RsaI</i> / <i>COL1A2</i>	$-0,077 \pm 0,084$	0,83	1	0,36
<i>MspI</i> / <i>COL1A2</i>	$0,185 \pm 0,184$	0,99	1	0,32
Media de los valores absolutos $\pm SE$	$0,076 \pm 0,018$			

En la tabla 54 se muestran los resultados del test χ^2 de contingencia para las frecuencias genotípicas del grupo con otosclerosis y del grupo control. En ningún caso se observa asociación significativa, los valores de probabilidad están muy alejados del valor de significación nominal ($P \gg 0,05$).

Tabla 54. Asociación de los genotipos en cada marcador con la otosclerosis

Marcador / Locus	χ^2	g.l.	P
<i>RsaI</i> / COL1A1	1,21	2	0,55
<i>MnII</i> / COL1A1	2,08	2	0,35
(GT) _n / COL1A2	8,54	20	0,99
<i>EcoRI</i> / COL1A2	1,77	2	0,41
(ACT) _n / COL1A2	10,78	14	0,70
<i>PvuII</i> / COL1A2	0,37	2	0,83
<i>RsaI</i> / COL1A2	0,87	2	0,65
<i>MspI</i> / COL1A2	1,07	2	0,59

Los análisis basados en haplotipos pueden detectar segmentos cromosómicos únicos que transporten alelos de predisposición a enfermedades porque cada nuevo alelo está asociado con su propia historia cromosómica (FALLIN *et al.* 2001). Además, evaluando haplotipos, más que realizando tests de asociación de un único locus, la pérdida de información atribuible a loci bialélicos sobre los multialélicos puede ser evitada y posiblemente mejorada (FALLIN *et al.* 2001; VARILO *et al.* 2003). Por lo tanto, se llevó a cabo una comparación de las frecuencias haplotípicas de dos loci entre la muestra de afectados y la muestra control.

Las frecuencias haplotípicas de dos marcadores se estimaron en casos y controles separadamente mediante el algoritmo E-M (expectación-maximización) (EXCOFFIER & SLATKIN 1995). En la tabla 55 se muestran los resultados de los tests χ^2 de homogeneidad realizados para las frecuencias haplotípicas de la muestra de

Resultados

otosclerosis y la muestra control, para cada uno de los pares de loci. Se observa que no existen diferencias estadísticamente significativas en ningún caso ($P > 0,05$).

Tabla 55. Asociación de haplotipos de dos marcadores con la otosclerosis

Par de marcadores / Locus	χ^2	g.l.	P
<i>RsaI</i> – <i>MnII</i> / COL1A1	2,17	3	0,54
(GT) _n – <i>EcoRI</i> / COL1A2	7,56	8	0,48
(GT) _n – (ACT) _n / COL1A2	17,25	7	0,44
(GT) _n – <i>PvuII</i> / COL1A2	12,27	10	0,27
(GT) _n – <i>RsaI</i> / COL1A2	6,39	8	0,60
(GT) _n – <i>MspI</i> / COL1A2	4,45	8	0,81
<i>EcoRI</i> – (ACT) _n / COL1A2	5,11	6	0,53
<i>EcoRI</i> – <i>PvuII</i> / COL1A2	2,10	3	0,55
<i>EcoRI</i> – <i>RsaI</i> / COL1A2	4,15	3	0,25
<i>EcoRI</i> – <i>MspI</i> / COL1A2	2,62	3	0,45
(ACT) _n – <i>PvuII</i> / COL1A2	17,25	17	0,44
(ACT) _n – <i>RsaI</i> / COL1A2	9,22	8	0,32
(ACT) _n – <i>MspI</i> / COL1A2	5,11	6	0,53
<i>PvuII</i> – <i>RsaI</i> / COL1A2	6,64	3	0,08
<i>PvuII</i> – <i>MspI</i> / COL1A2	1,04	2	0,60
<i>RsaI</i> – <i>MspI</i> / COL1A2	1,57	2	0,46

La otosclerosis presenta una prevalencia dos veces superior en mujeres que en hombres (BEN ARAB *et al.* 1993; GORDON 1989; MENDER & TANGE 2003). Por esta razón y para averiguar si el factor sexo está influenciando la manifestación de la enfermedad, se realizó el análisis de asociación de las

frecuencias alélicas por sexo. En la tabla 56 se muestran estos resultados y se puede observar que en ningún caso existen diferencias estadísticamente significativas ($P > 0,05$).

A la vista de todos los resultados anteriores se puede concluir que en nuestra población no se detecta asociación entre los genes del colágeno tipo I (*COL1A1* y *COL1A2*) y la otosclerosis.

Tabla 56. Análisis de asociación por sexos de alelos en cada marcador con la otosclerosis.

Marcador/Locus	Hombres			Mujeres		
	χ^2	g.l.	<i>P</i>	χ^2	g.l.	<i>P</i>
RsaI/COL1A1	2,16	1	0,14	0,40	1	0,53
MnII/COL1A1	0,19	1	0,67	2,10	1	0,15
(GT) _n /COL1A2	1,68	3	0,64	4,19	5	0,52
EcoRI/COL1A2	0,02	1	0,89	2,02	1	0,16
(ACT) _n /COL1A2	0,67	4	0,96	4,73	4	0,32
PvuII/COL1A2	0,08	1	0,78	0,05	1	0,83
RsaI/COL1A2	0,04	1	0,84	0,78	1	0,38
MspI/COL1A2	0,01	1	0,94	1,07	1	0,30

5. Discusión

EVALUACIÓN DEL DG DEPENDIENTE DE MARCADOR EN 11p

A pesar de la gran cantidad de estudios de DG de *background* existentes en la literatura, no se había tenido en cuenta hasta ahora el efecto del marcador en los niveles de DG detectados. Así, éste es el primer estudio que evalúa el DG dependiente de marcador, mediante la comparación de los patrones de DG de *background* entre pares de SNPs y entre pares de microsatélites, a lo largo de una región del genoma tan extensa como es el brazo corto del cromosoma 11 humano. El análisis se realizó a partir de dos muestras independientes con 405 individuos cada una, todos ellos provenientes de Galicia. El DG se evaluó entre todos los pares resultantes de la combinación de 18 SNPs que abarcan una región de 37,0 cM y 36,6 Mb, y entre todos los pares resultantes de la combinación de 22 microsatélites que se extienden a lo largo de 42,4 cM y 45,5 Mb. La FRE promedio ($\pm$ SE) entre todos los pares de SNPs y entre todos los pares de microsatélites analizados fue similar con valores de $18,3 \pm 1,0$ y $20,1 \pm 0,9$ cM, respectivamente. Así mismo, la DF promedio no presentó grandes diferencias, con valores de $16,1 \pm 1,1$ y $19,8 \pm 1,0$ Mb, entre pares de SNPs y entre pares de microsatélites respectivamente.

La evaluación del DG con cada tipo de marcador ofrece unos niveles y distribución diferentes del DG a lo largo de 11p. Así, un 6,5% de los pares de SNPs mostraron DG significativo, mientras que con microsatélites, el porcentaje de asociaciones significativas ascendió a un 16 % entre pares de alelos, y un 36 % entre pares de microsatélites. Estudios previos han sugerido que los marcadores multialélicos (microsatélites) ofrecen una mayor potencia estadística para detectar DG, frente a los marcadores bialélicos (SNPs) en condiciones equivalentes (OTT y RABINOWITZ 1997; CHAPMAN y WIJSMAN 1998; XIONG y JIN 1999; SHAM *et al.* 2000; VARILLO *et al.* 2003). Esta mayor potencia estadística es debida a una mayor heterozigosidad debido al mayor número de alelos de los loci microsatélites. Esta podría ser la causa de la mayor cantidad de DG observado en el presente trabajo entre pares de microsatélites respecto al DG entre pares de SNPs. Sin embargo,

Discusión

los datos sugieren la implicación de otros factores que parecen ser fundamentales y que se irán exponiendo a lo largo de esta discusión.

También se obtuvieron resultados diferentes en la intensidad del DG entre pares de SNPs y entre pares de microsatélites. Los pares de SNPs se asociaron entre sí no al azar con intensidades que oscilaron entre 0,13 y 0,99 y un valor promedio de $0,48 \pm 0,12$, medida en términos de $|D'|$. Entre pares de microsatélites las asociaciones interalélicas presentaron una intensidad mínima de 0,09, máxima de $\sim 1,0$, y un valor promedio de $0,34 \pm 0,01$, medida en términos de $D'_{ij}(+)$. Mientras que, la intensidad de las asociaciones globales entre pares de microsatélites osciló entre 0,05 y 0,54, con un valor promedio de $0,11 \pm 0,01$, medida en términos de $D'(+)_{ij}$.

Los fuertes contrastes en el comportamiento del DG evaluado entre pares de SNPs o entre pares de microsatélites, también se reflejaron en las diferentes FRE y DF observadas entre pares de marcadores en DG significativo. Las asociaciones significativas entre pares de SNPs incluyeron pares de loci tanto fuertemente (0 cM), como débilmente (19,8 cM) ligados. Sin embargo, la FRE promedio entre los pares significativos resultó $5,1 \pm 2,6$ cM, lo que indica una preponderancia de los pares de SNPs en DG separados por sólo unos pocos centimorgans. Así mismo, el rango de la DF entre pares de SNPs en desequilibrio osciló entre 0 y 16 Mb, con una media de $3,3 \pm 2,1$ Mb. Entre pares de microsatélites, los desequilibrios interalélicos significativos no estuvieron confinados a los pares más cercanos (0 cM y 0 Mb), sino que también estuvieron presentes entre los pares más distantes entre sí (42,4 cM y 45,5 Mb), con un promedio de $18,3 \pm 0,4$ cM y $18,0 \pm 0,4$ Mb. En lo que respecta a los desequilibrios globales, se obtuvieron asociaciones significativas desde 0 a 39,5 cM y desde 0 a 44,9 Mb, con promedios de $16,1 \pm 1,3$ cM y $15,5 \pm 1,4$ Mb, en función de la FRE y DF respectivamente.

Analizando los resultados con cada marcador separadamente, se encuentra que los bajos niveles de DG obtenidos entre pares de SNPs contrastan con los estudios previos. La región analizada en el presente trabajo entre pares de SNPs, alcanza los 37 cM y 37 Mb. Sin embargo, la mayoría de los trabajos estudiaron regiones pequeñas, de sólo unas pocas kilobases (BONNEN *et al.* 2000; EISENBARTH *et al.* 2000; REICH *et al.* 2001; KAESSMANN *et al.* 2002; TIRET *et al.* 2002; CLARK *et al.* 2003; VARILO *et al.* 2003; CARLSON *et al.* 2004), aunque algunos llegaron a estudiar pares de SNPs separados hasta por una o varias megabases (DUNNING *et al.* 2000; MARTIN *et al.* 2000; MOFFATT *et al.* 2000; TAILLON-MILLER *et al.* 2000; ABECASIS *et al.* 2001; EISENBARTH *et al.* 2001; SHIFMAN *et al.* 2003; KE *et al.* 2004; TSUNODA *et al.* 2004). Los resultados de estos trabajos informan, en general, de un elevado nivel de asociaciones no al azar entre SNPs separados por bajas frecuencias de recombinación o distancias físicas. MOFFATT *et al.* (2000) estudiaron 24 SNPs en 300 individuos de una población típicamente caucasoide. Los SNPs se distribuían a lo largo de 850 kb del brazo largo del cromosoma 14. Sus resultados reflejaron un 22 % (61/276) de asociaciones significativas, además, el DG era relativamente común a distancias de 250 kb, y todavía se detectaba por encima de 500 kb. En la misma línea, DUNNING *et al.* (2000) analizaron el DG entre pares de marcadores bialélicos en tres regiones cromosómicas diferentes, 1.335 kb en 13q12-13, 380 kb en 19q13.2 y 120 kb en 22q13.3. Los marcadores se analizaron en 1.600 individuos de cuatro poblaciones diferentes y los porcentajes de pares de SNPs en DG oscilaron entre el 23 % y el 36 %. En lo que respecta a la extensión del DG, estos autores no obtuvieron DG significativo a distancias mayores de 500 kb en ninguna de las poblaciones analizadas. TAILLON-MILLER *et al.* (2000) fueron los primeros en obtener DG a largas distancias entre pares de SNPs. Estos autores analizaron 39 SNPs en dos regiones cromosómicas, Xq25 y Xq28, para una muestra de población europea general (CEPH) y dos muestras de poblaciones aisladas de Finlandia y Cerdeña. Los resultados evidenciaron dos regiones extensas con un fuerte DG entre pares de SNPs, 1 Mb en Xq25 y 340 kb en Xq28. ABECASIS *et al.* (2001) examinando el DG en tres regiones cromosómicas con población británica, obtuvieron DG hasta una distancia de 50 kb en el 50 % de

Discusión

los pares de SNPs, además de asociaciones que ocasionalmente se extendían hasta 500 kb.

Si atendemos a los resultados del presente trabajo entre pares de loci microsatélites, se observa que los altos niveles de DG obtenidos a largas distancias son superiores a los habitualmente observados en los estudios recogidos en la literatura. (PETERSON *et al.* 1995; CROUAEU-ROY *et al.* 1996; TISHKOFF *et al.* 1996; LAAN y PÄÄBO 1997; TISHKOFF *et al.* 1998; HUTTLEY *et al.* 1999; EAVES *et al.* 2000; GORDON *et al.* 2000; TISHKOFF *et al.* 2000; WILSON y GOLDSTEIN 2000; ZAVATTARI *et al.* 2000; MOHLKE *et al.* 2001; SERVICE *et al.* 2001; HALL *et al.* 2002; KAESSMANN *et al.* 2002; RYBICKI *et al.* 2002; AULCHENKO *et al.* 2004). En estos trabajos se observa una baja proporción de pares de microsatélites en DG cuando éstos están separados por grandes FRE. PETERSON *et al.* (1995) realizaron el primer estudio de la distribución del DG entre pares de loci microsatélites a lo largo de una región extensa del genoma. La distancia máxima a la que obtuvieron DG fue superior a 18 cM, aunque la mayoría de las asociaciones estaban restringidas a los pares separados por menos de 4 cM, y el porcentaje total de pares de microsatélites asociados no al azar fue de tan sólo un 5 % (26/496). Los trabajos posteriores parecían confirmar el escaso alcance de las asociaciones entre pares de microsatélites en las poblaciones, de manera que la mayoría de las asociaciones significativas se extendían poco más de 1 cM. Así, HUTTLEY *et al.* (1999) estudiando población europea general obtuvieron tan solo un 4 % de microsatélites en DG separados por menos de 4 cM. Así mismo, en poblaciones que no han sufrido mezcla, WILSON y GOLDSTEIN (2000) no obtuvieron DG entre pares de microsatélites separados por más de 2 cM. Más recientemente, SERVICE *et al.* (2001), en una población de Costa Rica, analizaron 1036 microsatélites repartidos por todos los cromosomas y obtuvieron un 33 % de asociaciones significativas entre pares de microsatélites separados por menos de 4 cM.

La comparación de nuestros resultados con los estudios previos se ve obstaculizada por una serie de factores que difieren en unos y otros estudios y que

podrían tener un profundo efecto en los niveles de DG detectado. Entre estos factores se encuentran los métodos analíticos utilizados, el tamaño de muestra, el método de estimación de los haplotipos, la densidad de los marcadores a lo largo de la región analizada y el tamaño de la misma, y la historia demográfica de la población. En los trabajos presentes en la literatura, que describen el DG entre loci microsatélites a lo largo de regiones extensas de los cromosomas humanos, se estima únicamente el DG global. Al estudiar loci multialélicos, como los microsatélites, el desequilibrio global resume todos los desequilibrios existentes entre los pares de alelos de los loci en un único valor de intensidad y significación, con la consiguiente pérdida de información que esto conlleva, ya que, un par de loci estará o no en DG global independientemente de la cantidad e intensidad de los desequilibrios interalélicos entre el par de loci que se considera. Además, los alelos de loci multialélicos no pueden ser considerados como un conjunto homogéneo por lo que se necesita un análisis del DG dependiente de alelo. De esta manera, la medida de DG global no aporta información de los patrones de DG que puedan estar siendo influenciados por la frecuencia y tamaño de los alelos entre los pares de microsatélites (ZAPATA *et al.* 2001b).

En el presente estudio, la estimación de la cantidad de desequilibrios interalélicos y globales entre pares de microsatélites se realizó incluyendo únicamente los desequilibrios de signo positivo con el fin de obtener unos niveles más precisos de las asociaciones existentes en el brazo corto del cromosoma 11. Esta decisión vino motivada por la observación de que la potencia estadística del test chi-cuadrado es mayor para detectar desequilibrios positivos que negativos, cuando las clases haplotípicas en acoplamiento contienen los alelos más o menos frecuentes (BROWN 1975; THOMPSON *et al.* 1988; ZAPATA *et al.* 2001a, b). La consideración de los DG de signo positivo como indicadores de los niveles de DG en una región del genoma, no presenta ningún problema conceptual ya que, en principio, no existen razones que nos hagan pensar que en la naturaleza haya un porcentaje diferente de desequilibrios positivos que negativos. La mayor capacidad para detectar asociaciones positivas que negativas se debe a la asimetría de D_{max} cuando las frecuencias alélicas no son intermedias. En estas

condiciones, el valor que puede tomar el coeficiente D_{max} es siempre mayor para desequilibrios positivos que negativos. Además, la varianza de muestreo de las asociaciones negativas es mayor que las positivas cuando las frecuencias alélicas no son intermedias, por lo que podrían producir una sobreestima de la intensidad del desequilibrio (véase *Material y Métodos* y ZAPATA *et al.* 2001a, b). De hecho, la estimación del DG significativo entre pares de SNPs, no se realizó teniendo en cuenta el signo, dado que las frecuencias alélicas de los SNPs individuales fueron intermedias en casi todos los casos. Sólo 2 de los 18 SNPs presentaron una frecuencia relativa del alelo minoritario menor de 0,20 (DRD4-Smal y WT1-Hinfl). De estos dos SNPs, sólo el DRD4-Smal está implicado en un par en DG significativo (DRD4-Smal x HRASNsil). Además, el número de desequilibrios significativos positivos fue igual al de negativos, lo que parece indicar que estamos ante un caso en el que la potencia estadística es la misma para detectar desequilibrios de ambos signos.

Dada la importancia de los tamaños muestrales para disponer de una potencia estadística adecuada para detectar el DG, en el presente trabajo se utilizó un tamaño de muestra muy superior a algunos estudios previos (405 individuos) para evitar un rechazo espúreo de la hipótesis nula de DG. Así, la escasa detección de DG en muchos trabajos está asociada con una baja potencia estadística y con altas probabilidades de cometer un error de tipo II (probabilidad de rechazar la hipótesis nula siendo verdadera). Esta baja potencia estadística está causada por los pequeños tamaños muestrales manejados en muchos de los trabajos, cuando la intensidad del DG no es elevada y las frecuencias alélicas no son intermedias. Los tamaños de muestra requeridos para detectar (con un 90 % de probabilidad) el DG, dado el grado de polimorfismo de los marcadores utilizados, son substancialmente superiores a los habitualmente utilizados en la mayoría de los estudios (ZAPATA y ALVAREZ 1987, 1992, 1993; THOMPSON *et al.* 1988; ZERBA *et al.* 1991; FU y ARNOLD 1992).

Los estudios también difieren en los métodos de obtención de las frecuencias haplotípicas. En estudios en el cromosoma X los haplotipos se

obtienen directamente a partir de una muestra compuesta únicamente por hombres (TAILLON-MILLER *et al.* 2001; SANDE *et al.* 2003). Para datos de autosomas los haplotipos se pueden reconstruir en base a pedigríes, aunque éstos no siempre están disponibles. Por esta razón, en la mayoría de estudios con autosomas se estiman las frecuencias haplotípicas mediante métodos de iteración. En el presente trabajo se estimaron las frecuencias haplotípicas mediante métodos de iteración que emplean el algoritmo E-M (Expectación-Maximización) como herramienta (DEMPSTER *et al.* 1977; EXCOFFIER y SLATKIN 1995). Estudios previos habían obtenido diferentes frecuencias haplotípicas aplicando el algoritmo E-M que obteniendo los haplotipos a partir de datos familiares (MOFFATT *et al.* 2000). Pero, en general los estudios sugieren que el algoritmo E-M funciona bien bajo un amplio rango de escenarios poblacionales y de conjuntos de datos, incluso cuando algún alelo de los loci del par no cumple las proporciones H-W, para tamaños de muestra en el rango de moderado a grande (FALLIN y SCHORK 2000).

El test estadístico empleado para estimar la significación del DG puede tener una gran importancia en la capacidad de detección de las asociaciones no al azar. La baja potencia estadística de los tests de significación habitualmente utilizados para detectar el DG podría ser otro factor que estuviese influyendo en la escasa detección del DG en los estudios experimentales. Así, el test exacto de Fisher (FET) se utiliza en muchas ocasiones para testar independencia en tablas de contingencia 2 x 2. Pero se ha visto que posee un carácter extremadamente conservador, no deseado en la mayoría de los estudios. Así, el test chi-cuadrado resulta significativo más a menudo que el FET, lo que podría interpretarse como consecuencia de un bajo carácter conservador del test chi-cuadrado. Sin embargo, se ha visto que, incluso para bajos tamaños de muestra, el test X^2 convencional es robusto y generalmente proporciona un nivel de significación muy próximo o menor que el nivel de significación nominal (UPTON 1982; D'AGOSTINO *et al.* 1988; ZAPATA y ALVAREZ 1997). En el presente trabajo se ha aplicado el test chi-cuadrado con la corrección de Yates para obtener la significación del DG entre pares de SNPs y para obtener la significación del DG interalélico positivo entre pares de microsatélites. La significación del DG global positivo entre pares de

microsatélites se ha obtenido aplicando la corrección de Bonferroni a los tests chi-cuadrado individuales empleados para testar las asociaciones interalélicas dentro de cada par de microsatélites. La aplicación de esta metodología evita que se produzca un exceso de casos significativos como consecuencia del empleo de múltiples tests para comprobar la misma hipótesis (ZAPATA *et al.* 2001b). Sin embargo, esta técnica a la vez que disminuye la probabilidad de cometer un error de tipo I, provoca un aumento de la probabilidad del error de tipo II (ZERBA *et al.* 1991; ZAPATA y ALVAREZ 1993), por lo que proporciona una estima conservadora de la frecuencia del DG global positivo. De este modo, se esperaría que existiese una cantidad de DG global positivo significativo entre pares de microsatélites en 11p todavía mayor que la reflejada en este trabajo.

Un procedimiento utilizado frecuentemente en estudios de DG con loci multialélicos es el agrupamiento de los alelos con menor frecuencia en cada locus, en una única clase. Este procedimiento persigue convertir un sistema multialélico en dialélico y así reducir la complejidad de las tablas de contingencia $m \times n$, originadas en los sistemas multialélicos, a tablas 2×2 más fácilmente analizables. Este procedimiento se realiza cuando las frecuencias esperadas de las clases haplotípicas presentan un valor bajo. Sin embargo, puede producir una seria infraestima del DG presente en la población, ya que se reduce la potencia estadística de los tests habitualmente utilizados para testar el DG (WEIR y COCKERHAM 1978; WATKINS *et al.* 1994). Además, el agrupamiento de alelos oscurece gran cantidad de información que sería de utilidad para elucidar las fuerzas evolutivas que intervienen en la generación de los patrones del DG. Por último, la detección de patrones de DG dependientes de alelo se vería seriamente obstaculizada (ZAPATA *et al.* 2001b). Sin embargo, actualmente la teoría en loci multialélicos se encuentra lo suficientemente desarrollada como para que sea innecesaria la pérdida de información causada por el agrupamiento de alelos (ZAPATA *et al.* 2001b). En el presente trabajo no se realizó agrupamiento de alelos en ningún caso y únicamente se eliminaron del análisis los alelos de copia única o *singletons*, por los problemas que ocasionan tanto en la estimación de la frecuencia como de la intensidad del DG. El patrón de DG entre pares de

microsatélites se describió a nivel de los DG interalélicos positivos, lo que nos permitió realizar inferencias en el patrón de DG dependiente de alelo, y a nivel de los DG globales positivos. El DG entre pares de SNPs posee una interpretación más sencilla ya que, por su carácter bialélico, su análisis genera una única estima de DG global.

El tamaño de la región analizada podría explicar el bajo porcentaje de pares de SNPs en DG (6,5 %), respecto a los estudios previos. En el presente estudio se han analizado pares de SNPs con FRE y DF máximas de 37,0 cM y 36,6 Mb, respectivamente, muy superiores a las distancias máximas analizadas en trabajos anteriores con SNPs (PRITCHARD y PRZEWORSKY 2001). De hecho, si nos centramos exclusivamente en los pares de SNPs separados por menos de 1 cM o 1 Mb, los porcentajes de asociaciones no al azar ascienden visiblemente. Así, de los 153 pares de SNPs analizados, sólo 17 de ellos están separados por distancias menores de 1 cM y, de éstos, 7 pares presentan DG. Esto es, el 41 % de los pares de SNPs separados por menos de 1 cM presentan DG. Igualmente, si atendemos a la DF entre los pares de SNPs, el 67 % (8/12) de los pares separados por menos de 1 Mb están en DG. Estos porcentajes de pares de SNPs en DG a lo largo de 11p se asemejan más a los resultados recogidos en la literatura para otras regiones cromosómicas, e incluso son superiores a los observados en algunos casos. En este sentido cabe señalar que en un trabajo previo en el que se analizó una región menor, y por tanto menor número de SNPs, el porcentaje de pares en DG fue de un 14 % (9/66) (RODRÍGUEZ *et al.* 2003). En cambio, en el presente trabajo, al abarcar una porción mayor del genoma, aumentó mucho el número de pares de SNPs separados por grandes distancias que no mostraron DG significativo. En lo que respecta a los microsatélites, la densidad de los marcadores no parece ser una explicación probable para las diferencias observadas en la cantidad de DG entre éste y estudios previos, puesto que la FRE promedio para los pares de microsatélites en DG resultó muy superior en 11p (18 cM) a la reflejada en otros estudios.

Se ha sugerido que la historia demográfica poblacional puede estar influyendo en el DG detectado, de manera que las poblaciones de tamaño constante exhibirían mayores niveles de DG que las poblaciones que han experimentado una expansión reciente (SLATKIN 1994; CROUAEU-ROY *et al.* 1996; LAAN y PÄÄBO 1997; REICH *et al.* 2001; KAESMANN *et al.* 2002). Este fenómeno se comprobó experimentalmente en varios estudios que utilizaban loci microsatélites, así como en muchos otros trabajos con SNPs. Así, LAAN y PÄÄBO (1997) analizando el DG a lo largo del cromosoma X con microsatélites, obtuvieron DG a mayores distancias (hasta 6 cM) en una población de tamaño constante, respecto a tres poblaciones de reciente expansión donde apenas se detectaba DG entre loci próximos. Más recientemente, REICH *et al.* (2001) analizando el DG entre pares de SNPs en 19 regiones del genoma, observaron que la extensión del DG era dependiente de la historia demográfica de la población analizada. Sus resultados evidenciaron una mayor extensión del DG en poblaciones que han sufrido un cuello de botella en el pasado, como poblaciones del Norte europeo, frente a poblaciones como las del África subsahariana. Además, otros trabajos informan de diferencias poblacionales en la proporción y extensión del DG, tanto entre pares de SNPs, como entre pares de microsatélites, con una mayor proporción de pares de marcadores en desequilibrio separados por grandes distancias en poblaciones aisladas respecto a la población europea general (GODDARD *et al.* 2000; KIDD *et al.* 2000; KAESMANN *et al.* 2002; VARILLO *et al.* 2003). Así, KAESMANN *et al.* (2002) estudiando el DG entre 50 SNPs en cinco genes diferentes, obtuvieron unos porcentajes de DG del 60 % y del 48 % en dos poblaciones estables y aisladas históricamente, y del 29 % en las poblaciones sueca y finlandesa generales que han experimentado un proceso de expansión reciente. Para estimar el DG a mayores distancias, estos autores analizaron 14 microsatélites separados por distancias comprendidas entre 1 y 16 cM en el cromosoma X. En poblaciones aisladas obtuvieron un 22 % y un 13 % de asociaciones significativas que se extendían hasta 4 cM, mientras que, en las poblaciones sueca y finlandesa generales el porcentaje fue tan solo del 1 % y la distancia máxima a la que se detectaba DG, tan solo de 0,5 cM. Nuestros resultados mostraron que el porcentaje de pares de SNPs separados una

distancia ≤ 1 cM, así como de pares de microsatélites, con DG significativo, es similar al hallado en los trabajos anteriormente citados para poblaciones aisladas con una historia demográfica estable. No obstante, la historia demográfica parece una explicación poco probable para la generación del elevado porcentaje observado de asociaciones significativas. De hecho, las evidencias disponibles acerca de la historia demográfica de la población gallega, son compatibles con un modelo de expansión poblacional en Europa durante el Paleolítico Superior (13.000 – 26.000 años) coincidiendo con la irrupción de la agricultura (CAVALLI-SFORZA *et al.* 1994; SALAS *et al.* 1998).

Una mayor dificultad estriba en la comparación de la intensidad del DG global entre loci microsatélites obtenida en diferentes trabajos. Muchos autores han medido la intensidad del DG global entre pares de microsatélites mediante las probabilidades resultantes de los tests de significación. Así, un par de loci se encontrará más fuertemente asociado cuanto mayor sea el valor de la probabilidad resultante del test de significación de la muestra gamética. Sin embargo, ésta no es una medida válida de la intensidad de desequilibrio entre un par de marcadores para realizar comparaciones entre diferentes poblaciones y diferentes regiones cromosómicas, puesto que su valor depende del tamaño de muestra, de las frecuencias alélicas y del signo de las asociaciones (BROWN 1975; WEIR y COCKERHAM 1978; THOMPSON *et al.* 1988; OTT y RABINOWITZ 1997; ZAPATA *et al.* 2001b). Así, la mejor manera de medir la intensidad del DG son los coeficientes o índices de asociación (ZAPATA *et al.* 2001b). En el presente trabajo, la intensidad del DG global entre pares de microsatélites se estimó mediante el coeficiente $D'(+)$, que se calcula únicamente a partir de los valores de los desequilibrios interalélicos de signo positivo. Al no considerar los desequilibrios de signo negativo, el coeficiente $D'(+)$ posee una menor varianza de muestreo y, por tanto, ofrece una estima más fiable y precisa de la magnitud del DG global entre loci microsatélites (ZAPATA *et al.* 2001b). Los valores de la magnitud del DG global, medido en términos de $D'(+)$, se distribuyeron a lo largo de un rango amplio, desde un mínimo de 0,05 hasta un máximo de 0,54. A pesar de este rango tan amplio, la mayoría de los casos no superaron el valor de 0,25. Así, el 99 % (83/84) de los

Discusión

pares de microsatélites en DG significativo presentaron valores de $D'(+)$ inferiores a 0,25. La baja magnitud de las asociaciones entre pares de microsatélites también se refleja en su media, que informa de que el DG global entre pares de microsatélites es de baja intensidad ($0,11 \pm 0,01$). Este bajo nivel de intensidad del DG global entre pares de microsatélites confirma resultados previos, obtenidos empleando la misma metodología de estimación, con 11 microsatélites distribuidos a lo largo del brazo largo del cromosoma X humano (Xq25-28) (SANDE 2002; SANDE *et al.* 2003), así como en la región cromosómica 11p15 (ZAPATA *et al.* 2001b).

Los resultados del DG obtenidos entre pares de SNPs contrastan fuertemente con los valores de DG global obtenidos con loci microsatélites. Con un porcentaje de pares de SNPs en DG de un 6,5 % (10/153), no todos los SNPs participan en asociaciones no al azar. Así, de los 18 SNPs analizados sólo 11 están formando parte de alguna asociación significativa y, entre ellos, el SNP H19Alul está implicado en el 40 % (4/10) de todas las asociaciones. La magnitud de las asociaciones no al azar, medida en términos de $|D'|$, abarca casi todo el rango posible de valores, desde el DG de baja intensidad (0,13, par H19Alul x TpHDrall), hasta un valor de alta intensidad de DG próximo a 1,00 (0,99, par INSPstl x INSDrall). El 40 % (4/10) de los desequilibrios significativos son de elevada intensidad, con un valor de $|D'| > 0,5$. Así, los SNPs se asociaron no al azar con una intensidad promedio de $0,48 \pm 0,12$.

El estudio de las asociaciones interalélicas positivas entre pares de microsatélites llevado a cabo en el presente trabajo, nos permitió una caracterización más adecuada de los patrones de DG entre pares de loci microsatélites a lo largo de 11p. El porcentaje de pares de alelos que mostraron DG positivo significativo resultó del 16% (1211/7512). Los desequilibrios interalélicos positivos se distribuyeron por casi toda la región analizada y sólo un 2 % (4/231) de los pares de microsatélites no presentaron ninguna asociación interalélica positiva significativa (D11S4185 x D11S1323, D11S1360 x D11S4177, D11S1763 x D11S1331, D11S903 x D11S4185). En general, el porcentaje de

asociaciones interalélicas positivas significativas fue bastante heterogéneo entre todos los pares de microsatélites. Así, el porcentaje de asociaciones interalélicas positivas significativas por par de microsatélites osciló entre un 4 % y un 50 %, con un valor promedio de $16,3 \pm 0,5$ %. Los valores de $D'_{ij}(+)$ entre los pares de alelos de loci microsatélites cubrieron casi todo el rango posible, desde 0,09 a 1,00, con un valor promedio de $0,34 \pm 0,01$. De esta manera, estos datos reflejan una intensidad promedio moderada del DG interalélico para todos los pares de microsatélites. Así mismo, la figura 13 del apartado de *Resultados* muestra una preponderancia de desequilibrios interalélicos de intensidad moderada para la mayoría de los casos. A modo ilustrativo véase que un 70 % (853/1211) de los DG interalélicos significativos poseen valores de $D'_{ij}(+)$ inferiores a 0,40. Sirvan estas observaciones como contraste a la mayor intensidad del DG entre pares de SNPs, donde la magnitud promedio resultó próxima al 50 % de su máximo valor posible ($|\overline{D}| = 0,48 \pm 0,12$).

Las diferencias observadas, tanto en la frecuencia como en la intensidad del DG entre pares de SNPs y entre pares de microsatélites, indican que el DG dependiente de marcador determina una parte importante del DG de *background* observado en 11p. Los estudios recogidos en la literatura no habían tenido en cuenta el efecto del marcador en los niveles de DG observados, de manera que cruzaban información obtenida con distintos tipos de marcadores. Sin embargo, en el presente estudio se observa que los resultados obtenidos con diferentes marcadores no son necesariamente comparables.

Una variedad de agentes evolutivos, además de la acción de la selección, pueden ser los responsables de la generación y mantenimiento en las poblaciones de los patrones de DG a lo largo del genoma, como deriva génica o efectos fundadores (HILL y ROBERTSON 1968; AVERY y HILL 1979), estratificación o mezcla de poblaciones (CHAKRABORTY y WEISS 1988; WILSON y GOLDSTEIN 2000) y sucesos mutacionales recientes (ZAPATA *et al.* 2001). Además, alguno de estos

factores podría estar operando de distinta manera en SNPs y en microsatélites para provocar los diferentes patrones observados.

El papel de la selección natural en la generación o mantenimiento de los patrones de DG es difícil de evaluar. En el presente trabajo, los 18 SNPs analizados se encuentran en áreas no codificadoras del genoma, tanto en regiones intergénicas como formando parte de intrones en genes codificadores. Así, en principio, estos SNPs no estarían sometidos a la acción de la selección natural. Igualmente, los microsatélites se localizan en regiones no codificadoras del genoma y se han considerado neutrales a la acción de la selección natural. Sin embargo, existen situaciones en las que la selección puede generar DG entre loci neutrales, en áreas localizadas del genoma (HUTTLEY *et al.* 1999; PRZEWORSKI 2002). Así, si un locus que está bajo la acción de la selección natural está ligado a otro neutral, éste modificará sus frecuencias alélicas en un grado que dependerá de la intensidad de la selección y de la frecuencia de recombinación entre ambos (MAYNARD SMITH y HAIGH 1974; NACHMAN 2001). Este fenómeno de arrastre genético o *hitchhiking* puede generar asociaciones nuevas o romper otras preexistentes entre los loci. Se distinguen dos fenómenos de arrastre genético: El primero, barrido selectivo o *selective sweeps*, ocurre cuando la selección natural actúa sobre mutaciones ventajosas (KOJIMA y SCHAEFFER 1967; MAYNARD SMITH y HAIGH 1974; OHTA y KIMURA 1975; KAPLAN *et al.* 1989; WIEHE 1998; NACHMAN 2001; PRZEWORSKI 2002) y el segundo, denominado selección de fondo o *background selection*, está basado en la selección actuando en contra de mutaciones deletéreas (CHARLESWORTH *et al.* 1993, 1995; NACHMAN 2001). En el presente trabajo, el DG detectado podría ser debido al efecto de arrastre genético generado por genes ligados sometidos a selección, aunque es poco probable que este efecto se extienda a lo largo de toda la región analizada. Además, la ausencia de tests específicos para discriminar entre los orígenes selectivos y neutrales del DG a una escala genómica, no permite una conclusión formal respecto a la implicación de la selección natural en el origen del DG detectado (HUTTLEY *et al.* 1999). Por otro lado, la selección no explica las diferencias en los niveles de DG obtenidos entre pares de SNPs y entre pares de microsatélites, puesto que los dos

tipos de marcadores se localizan a lo largo de la misma región cromosómica y, en principio, estarían sometidos a las mismas fuerzas selectivas. Sin embargo, la mayor tasa mutacional de los microsatélites podría restablecer su variabilidad relativamente rápido haciéndolos menos sensibles al fenómeno de *hitchhiking* (WIEHE 1998). De esta manera, podría ocurrir que la selección natural esté actuando en esta región modelando los niveles de DG, aunque otros factores evolutivos parecen estar implicados en la generación y mantenimiento de los diferentes patrones de DG entre SNPs y entre microsatélites en el brazo corto del cromosoma 11.

La deriva genética se ha estudiado en múltiples ocasiones al respecto de la generación de diferencias en las frecuencias haplotípicas y su concomitante variación en el DG (OHTA 1982; SVED 1971). Se ha observado que la deriva genética representa un importante papel en la generación de DG en poblaciones aisladas (SLATKIN 1994; LAAN y PÄÄBO 1997). Así mismo, este factor se encuentra fuertemente inhibido en poblaciones que han sufrido una expansión reciente (SLATKIN 1994). Por esta razón, es poco probable que este factor esté generando el patrón de DG observado en el presente trabajo puesto que la historia demográfica de la población gallega es compatible con un modelo de expansión poblacional en Europa durante el Paleolítico Superior (SALAS *et al.* 1998). Además, la deriva genética actúa a escala genómica, de manera que su efecto no crea diferencias entre regiones cromosómicas o entre marcadores genéticos distintos. Por lo tanto, los diferentes niveles de DG observados entre pares de SNPs y entre pares de microsatélites no se justifican por la acción de la deriva.

La mezcla reciente de poblaciones genéticamente diferenciadas es un factor que puede influir en la generación de DG. En el momento de la mezcla, puede existir DG incluso entre sitios no ligados (CHAKRABORTY y WEISS 1988; PRITCHARD y ROSENBERG 1999). Sin embargo, éste decae muy rápidamente: con apareamiento aleatorio, el DG decae con una tasa de $(1 - c)$ por generación (HARTL y CLARK 1997). El resultado neto es que en poblaciones con una mezcla reciente, puede existir DG a largas distancias (WILSON y GOLDSTEIN 2000; SERVICE

et al. 2001; RYBICKI *et al.* 2002). Mediante el estudio de las desviaciones de las proporciones H-W podemos detectar la existencia de subheterogeneidad poblacional, ya que la mezcla de grupos poblacionales con diferentes frecuencias alélicas genera un déficit de heterocigotos. Sin embargo, es poco probable que en el presente estudio exista una estratificación poblacional subyacente que esté generando el DG observado, ya que las frecuencias genotípicas de los 18 SNPs y 22 microsatélites analizados están de acuerdo con las proporciones H-W. Aunque en una sola generación de apareamiento aleatorio se pueden alcanzar las frecuencias genotípicas esperadas bajo equilibrio H-W, la mezcla de poblaciones en el pasado podría haber generado un DG elevado a largas distancias durante muchas generaciones (CHAKRABORTY y WEISS 1988; STEPHENS *et al.* 1994). Sin embargo, Galicia es una región que ha permanecido relativamente aislada del resto de la península, con altas tasas de emigración y sin apenas inmigración (SALAS *et al.* 1998).

Nuestros análisis sugieren que la distinta dinámica mutacional de SNPs y microsatélites podría ser el factor determinante para las diferencias observadas en los patrones de DG entre pares de SNPs y entre pares de microsatélites (DG dependiente de marcador) (BOEHKE 2000; JORDE 2000; VARILLO *et al.* 2003). ZAPATA *et al.* (2001) propusieron que el DG observado entre pares de microsatélites podía deberse a la dinámica mutacional de los microsatélites en la región telomérica del brazo corto del cromosoma 11. El análisis de los patrones del DG dependiente de la frecuencia y tamaño de los alelos, nos ofrece una vía para interpretar las asociaciones observadas entre pares de microsatélites en 11p. Así, nuestros resultados parecen confirmar estos estudios previos, donde se sugiere que los desequilibrios entre alelos en baja frecuencia podrían tener una aportación importante al patrón de DG entre pares de microsatélites (ZAPATA *et al.* 2001b; SANDE *et al.* 2003). En el presente trabajo se realizó un análisis del DG entre pares de microsatélites, eliminando de la muestra de haplotipos aquellos que incluían alelos con una frecuencia menor al 3 %. De esta manera se obtuvo un porcentaje de desequilibrios interalélicos positivos significativos entre los alelos más frecuentes del 13 % (449/3417). Al estratificar la muestra de haplotipos en

haplotipos de dos alelos raros (RR), haplotipos de un alelo raro y uno común (RC) y haplotipos de dos alelos comunes (CC), se observaron diferencias estadísticamente significativas en la cantidad de asociaciones en cada clase. Así, los porcentajes de asociaciones interalélicas significativas fueron del 52 % (221/428), 15 % (541/3667) y 13 % (449/3417), para los grupos de haplotipos RR, RC y CC, respectivamente. Además, estas diferencias no se pueden atribuir a diferentes FRE ya que los promedios de éstas resultaron muy similares en los diferentes grupos de haplotipos ($19 \pm 0,9$, $18 \pm 0,5$ y $19 \pm 0,6$, respectivamente). Este resultado debería sorprendernos ya que habitualmente la potencia estadística para detectar DG es menor en el caso de frecuencias alélicas extremas (BROWN 1975; ZAPATA y ALVAREZ 1992, 1993). Además, podría argumentarse que alguno de los desequilibrios detectados entre alelos en baja frecuencia fuesen artefactuales, debido a que, cuando los valores esperados son bajos, el test χ^2 en tablas de contingencia 2 x 2 puede ser poco conservativo y rechazar la hipótesis nula con demasiada frecuencia. Sin embargo, se ha visto que el test χ^2 es robusto y generalmente proporciona un nivel de significación cercano o menor al nivel de significación de 0,05 (UPTON 1982; D'AGOSTINO *et al.* 1988; ZAPATA y ALVAREZ 1997). Además, para minimizar los efectos de posibles falsos positivos, se aplicó el test χ^2 con la corrección de Yates, que se considera conservativa (UPTON 1982; D'AGOSTINO *et al.* 1988). Es de destacar que la magnitud máxima de asociación no al azar entre los pares de alelos más frecuentes fue de 0,77, lo cual contrasta con el valor de asociación máxima considerando todos los alelos (1,00). Así mismo, el valor medio de la intensidad de las asociaciones interalélicas, medida en términos de $D'_{ij}(+)$, resultó claramente inferior en la muestra de los alelos más frecuentes ($0,21 \pm 0,01$) respecto al valor medio considerando todos los alelos ($0,34 \pm 0,01$). De esta manera, los resultados sugieren que los alelos que están contribuyendo a los desequilibrios de alta intensidad son los alelos menos frecuentes.

Igualmente, el DG parece estar relacionado con el tamaño de los alelos implicados (ZAPATA *et al.* 2001b). En el presente trabajo, los resultados mostraron una cantidad significativamente superior de desequilibrios interalélicos entre haplotipos formados por dos alelos de tamaño extremo (EE, 55 % de pares

significativos en esta clase), respecto a haplotipos formados por un alelo de tamaño extremo y un alelo de tamaño intermedio (EI, 17 % de pares significativos en esta clase), y respecto a haplotipos formados por dos alelos de tamaño intermedio (II, 14 % de pares significativos).

Los alelos en baja frecuencia y de tamaño extremo tienden a presentar DG mucho más a menudo que los alelos más comunes y de tamaño intermedio. Este hecho se podría interpretar como consecuencia de la elevada tasa mutacional (10^{-5} – 10^{-3}) de los microsatélites (CHAKRABORTY *et al.* 1997; WEBER y WONG 1993; BRINKMAN *et al.* 1998; KRUGLYAK *et al.* 1998; PRIMMER *et al.* 1998; XU *et al.* 2000). Así muchos de los desequilibrios detectados podrían ser una consecuencia de eventos mutacionales relativamente recientes. De hecho, cuando un alelo nuevo aparece en la población por efecto de una mutación, éste surge en DG completo con los alelos del haplotipo donde se originó. Con el paso de las generaciones, estas asociaciones se van rompiendo mientras la recombinación tiene lugar. Sin embargo, debido a la elevada tasa mutacional de los microsatélites, continuamente están surgiendo alelos nuevos y generando asociaciones, de manera que la recombinación no tiene tiempo de aleatorizar esas asociaciones en los gametos. Según la teoría genético-poblacional, la frecuencia alélica se relaciona con la antigüedad de ese alelo (considerado neutral) en la población (WATERSON y GUESS 1977). Los alelos en baja frecuencia tienen una mayor probabilidad de haber surgido recientemente en la población, respecto a los alelos de frecuencia moderada, puesto que es necesario el paso de las generaciones para que un alelo llegue a ser frecuente (KRUGLYAK 1999). Indudablemente, el decaimiento del DG entre loci distantes depende principalmente de la fracción de recombinación y del número de generaciones transcurridas desde que una mutación tuvo lugar. Así, el DG generado por mutaciones recientemente introducidas, lleva en la población pocas generaciones y su dilución debe estar todavía en progreso.

El patrón de DG entre pares de microsatélites dependiente del tamaño de los alelos, también presenta una explicación basada en la dinámica mutacional de

los microsatélites (ZAPATA *et al.* 2001b). Así, las constricciones, selectivas o no, en el número de repeticiones del dinucleótido, dictan mayores tasas de conversión hacia alelos de tamaño extremo. De esta forma, estas nuevas variantes alélicas de tamaño extremo estarían surgiendo continuamente en DG, lo que es consistente con la observación de que estos alelos ocurren a muy baja frecuencia, tanto en esta población como en otras (VALDES *et al.* 1993; SCHLÖTTERER *et al.* 1998; WEI *et al.* 1999; ZAPATA *et al.* 2001b). Por el contrario, los alelos comunes tienen una mayor probabilidad de ser más antiguos en la población, son generalmente de tamaño intermedio y, por lo tanto, están sujetos a menos constricciones y sufren un intercambio menos frecuente. De esta manera, dada la estabilidad de los alelos comunes y el tiempo transcurrido desde la introducción de las asociaciones, la recombinación será más eficiente rompiendo el DG entre ellos. En este sentido, la persistencia de algunos desequilibrios entre alelos comunes podrían reflejar los niveles de DG de *background* en 11p (ZAPATA *et al.* 2001b).

La tasa de mutación para SNPs se estima en torno a 1×10^{-9} por generación (LI *et al.* 1981; LI 1997; MARTINEZ-ARIAS *et al.* 2001). Por esta razón, la probabilidad de que ocurran dos cambios nucleotídicos independientes en el mismo sitio, así como de una mutación recurrente, es muy baja. De esta manera, en un SNP bialélico es muy improbable que surjan nuevos alelos por mutación que generen nuevas asociaciones y alteren el patrón existente de DG (XIONG y JIN 1999). Así, las asociaciones entre pares de SNPs tendrán una mayor antigüedad en la población que las asociaciones entre pares de alelos de microsatélites, por lo que la recombinación habrá tenido más tiempo de actuar rompiendo las asociaciones entre pares de SNPs alejados en el cromosoma. Entonces, los SNPs retendrán la historia demográfica más tiempo que los microsatélites, debido a su menor tasa de mutación (VARILLO *et al.* 2003), por lo que serán mejores indicadores de fenómenos evolutivos ocurridos en el pasado.

Los análisis de DG en humanos han demostrado ser extremadamente útiles para la localización de genes implicados en enfermedades de herencia mendeliana. Por el contrario, su utilidad en estudios de asociación a lo largo de

todo el genoma, propuestos como un medio para identificar los numerosos genes de débil efecto implicados en la susceptibilidad de enfermedades complejas, permanece siendo un tema controvertido (RISCH y MERIKANGAS 1996; JORDE 2000; RISCH 2000; ALTMÜLLER *et al.* 2001; LOHMUELLER *et al.* 2003; NEALE y SHAM 2004; ZONDERVAN y CARDON 2004). La existencia de niveles útiles de DG a largas distancias es un prerequisite que reduciría el número de marcadores requeridos para realizar un estudio de asociación en todo el genoma (COLLINS *et al.* 1999; KRUGLYAK 1999; BOEHNKE 2000; OTT 2000). En un primer momento, las predicciones teóricas fueron muy pesimistas sugiriendo que el DG útil entre pares de SNPs no se extendería más allá de 3 kb en la población general (KRUGLYAK 1999). De esta manera, se necesitarían 500.000 SNPs repartidos por todo el genoma para realizar un estudio de asociación. Pero en estudios empíricos posteriores se observó que el DG entre pares de SNPs se extendía a mayores distancias, entre 50 kb y 1 Mb (JORDE *et al.* 1994; BOEHNKE 2000; DUNNING *et al.* 2000; TAILLON-MILLER *et al.* 2000; ABECASIS *et al.* 2001; REICH *et al.* 2001; CLARK *et al.* 2003; VARILLO *et al.* 2003). Sin embargo, el DG entre pares de microsatélites se extendía hasta FRE máximas de unos pocos a 19 cM (PETERSON *et al.* 1995; EAVES *et al.* 2000; GORDON *et al.* 2000; WILSON y GOLDSTEIN 2000; SERVICE *et al.* 2001; ZAPATA *et al.* 2001b; KAESMANN *et al.* 2002). En el presente trabajo, se observa que el efecto del marcador es decisivo para indicar una distancia máxima de detección del DG. Así, se ha observado DG entre pares de SNPs separados por distancias máximas de 20 cM, aunque la mayoría de las asociaciones significativas (70 %) no se extendieron más allá de 0,5 cM. De manera que la FRE promedio entre los pares de SNPs en DG resultó de $5 \pm 2,6$ cM, incluso este valor es muy elevado por el efecto que ejercen en el promedio los dos pares de SNPs en DG separados por 20 cM. Estos resultados contrastan fuertemente con los obtenidos entre pares de microsatélites, ya que el DG se detectó a lo largo de toda la región analizada tanto entre los pares más cercanos (0 cM), como entre los más separados entre sí (42 cM), con un promedio de $18 \pm 0,4$ cM. Este resultado es sorprendente ya que se obtiene DG a una distancia que roza la segregación independiente. La mayor tasa mutacional de los microsatélites, parece ser la principal responsable del diferente comportamiento de ambos tipos de marcadores

frente a la acción aleatorizadora de la recombinación. Así, las asociaciones entre pares de SNPs llevarán más tiempo en la población, debido a la antigüedad de los alelos implicados, de manera que la recombinación lleva tiempo aleatorizando las asociaciones entre SNPs alejados. En los loci microsatélites, debido a la existencia de mutaciones frecuentes, continuamente están apareciendo alelos nuevos que, en un primer momento, surgen en DG completo, incluso con loci situados a largas distancias; de manera que a la recombinación no le da tiempo de romper todas las asociaciones nuevas que surgen entre loci microsatélites. Sin embargo, la mayor tasa de mutación de los microsatélites podría provocar un DG menor del esperado a cortas distancias, donde la tasa de recombinación puede verse superada por la tasa de mutación (LITT y JORDE 1986; XIONG y JIN 1999). Por el contrario, algunos estudios concluyeron que los microsatélites podrían ser unos marcadores adecuados para detectar DG a cortas distancias (SCHULZE *et al.* 2002). De hecho, este efecto no se observó en nuestros resultados, puesto que se detectó DG significativo entre pares de microsatélites separados tanto por FRE bajas (0 cM), como elevadas (42 cM).

El presente estudio es una extensión del trabajo previo realizado por ZAPATA *et al.* (2001b) con microsatélites a lo largo de la región más telomérica del brazo corto del cromosoma 11. Estos autores obtuvieron DG entre pares de microsatélites a lo largo de toda la región analizada (19 cM), de manera que no determinaron cuál era la distancia máxima a la que se podía detectar DG entre pares de microsatélites. En nuestro caso, la región analizada abarca todo el brazo cromosómico (11p) y, sin embargo, tampoco se ha logrado establecer una distancia máxima a partir de la cual no detectaríamos DG entre pares de microsatélites. Es importante señalar que se detectó DG a una distancia que roza la segregación independiente, por lo que no parece existir límite para la existencia de asociaciones no al azar entre pares de microsatélites.

Nuestros resultados sugieren que la distancia máxima a la que se detecta DG varía en función del marcador genético. Entre pares de SNPs se detecta DG a distancias muy inferiores a las observadas entre pares de microsatélites. Este

Discusión

patrón de DG dependiente de marcador también se puede observar con el estudio de otro tipo de marcadores genéticos. Así, por ejemplo, en *D. melanogaster* se ha observado DG entre pares de loci proteicos relacionados funcionalmente separados por distancias muy considerables (FRE media de 9 cM). En este caso los resultados sugerían la presencia de interacciones epistáticas muy fuertes entre los pares de loci proteicos (ZAPATA *et al.* 2002).

En base a la teoría genética poblacional clásica, el DG decaería monótonamente al aumentar la distancia entre los marcadores genéticos, reflejando de esta manera la acción de la recombinación a través de las generaciones pasadas. Pero este patrón puede verse distorsionado dada la gran cantidad de factores que pueden actuar junto con la recombinación, como selección natural, mezcla de poblaciones, deriva genética, mutación y conversión génica actuando a cortas distancias (LITT y JORDE 1986; WATKINS *et al.* 1994; PRITCHARD y PRZEWORSKI 2001). En principio se esperaría que el DG se disipase más rápidamente en las regiones teloméricas que en las pericentroméricas, ya que la tasa de recombinación parece ser mayor en las regiones teloméricas de los cromosomas humanos (BUETOW *et al.* 1991; WATKINS *et al.* 1994; TIRET *et al.* 2002). Sin embargo, en el presente trabajo no se han detectado diferencias apreciables en los niveles de DG entre la región 11p15 de 19 cM, más telomérica (18 % de asociaciones interalélicas significativas), y el resto del brazo 11p de 23 cM (19 % de asociaciones interalélicas significativas).

Los resultados del presente trabajo reflejaron una cantidad significativamente superior de asociaciones entre pares de SNPs separados por distancias ≤ 2 cM o 2 Mb respecto a los pares separados por distancias > 2 cM o 2 Mb ($P < 0,001$). Entre pares de microsatélites, también el número de asociaciones interalélicas resultó significativamente superior entre los pares de alelos separados por menos de 2 cM o 2 Mb, respecto a los pares de alelos más alejados (> 2 cM o 2 Mb) ($P < 0,05$). Estos resultados están de acuerdo con los estudios previos, donde se informa de una mayor cantidad de DG entre pares de loci estrechamente

ligados (PETERSON *et al.* 1995; HUTTLEY *et al.* 1999; EAVES *et al.* 2000; DUNNING *et al.* 2000; TAILLON-MILLER *et al.* 2000).

El estudio empírico de la relación de la intensidad del DG con la FRE o DF ha obtenido resultados diversos, tanto con SNPs como con microsatélites (JORDE *et al.* 1994; PETERSON *et al.* 1995; HUTTLEY *et al.* 1999; DUNNING *et al.* 2000; TAILLON-MILLER *et al.* 2000; WILSON y GOLDSTEIN 2000; DALY *et al.* 2001; JEFFREYS *et al.* 2001; NAKAJIMA *et al.* 2002). Así, DUNNING *et al.* (2000) obtuvieron una correlación negativa del DG entre pares de SNPs con la DF, que se extendía hasta 500 kb. Igualmente, PETERSON *et al.* (1995) y HUTTLEY *et al.* (1999), entre otros, estudiando microsatélites, obtuvieron una correlación negativa entre el DG y la FRE entre pares de loci. Sin embargo, otros estudios, como el de TAILLON-MILLER *et al.* (2000) y el de EAVES *et al.* (2000) no observaron una tendencia a disminuir la intensidad de las asociaciones entre pares de SNPs y entre pares de microsatélites respectivamente, al aumentar la DF entre ellos.

La dinámica evolutiva de cada tipo de marcador puede ser fundamental para el mantenimiento o no de una tendencia monótona de decaimiento de la intensidad del DG con una mayor FRE o DF. Nuestros análisis a lo largo de 11p revelaron una correlación negativa significativa entre la intensidad del DG y la FRE o DF entre pares de SNPs ($P < 0,01$). Así, el DG entre pares de SNPs tiende a disminuir al aumentar la distancia entre ellos, lo que implica que la recombinación entre pares de SNPs está ejerciendo un efecto preponderante frente a otros fenómenos evolutivos que podrían alterar esa tendencia. Por el contrario, no se detectó relación de la intensidad del DG interalélico con la FRE ni con la DF entre pares de microsatélites. Esta diferencia parece ser debida, de nuevo, a la mayor tasa mutacional de los loci microsatélites. De hecho, la correlación de la intensidad del DG con la FRE, eliminando de la muestra los alelos menos frecuentes ($p_i, q_j < 3\%$), resultó significativa ($P < 0,05$). Esta tendencia se puede explicar por la diferente edad de los desequilibrios, de manera que los desequilibrios entre los alelos más comunes, y posiblemente más antiguos en la población, habrán estado expuestos durante un período de tiempo mayor a la acción de la recombinación,

que los desequilibrios entre los alelos más recientes. La observación de que los alelos más recientes en la población se escapan a la acción aleatorizadora de la recombinación, ya fue señalada por ZAPATA *et al.* (2001b) para la región más telomérica incluida en el presente trabajo (19 cM). Estos autores señalaron, además, que la fuerza de la correlación negativa del DG interalélico con la FRE o DF entre pares de microsatélites, aumentaba al considerar alelos cada vez más comunes. En cuanto a los SNPs, su estabilidad mutacional, junto con el hecho de haber utilizado una mayoría de SNPs comunes (sólo dos de los 18 SNPs presentaron una frecuencia del alelo minoritario menor de 0,2), implicaría que los desequilibrios entre pares de SNPs son antiguos y llevan tiempo en la población sometidos a la acción de la recombinación.

El mantenimiento de una correlación negativa entre el DG y la FRE o DF es un punto importante para poder realizar estudios de clonaje posicional a largas distancias mediante DG con marcadores genéticos. Nuestros resultados sugieren que los SNPs son unos marcadores apropiados para realizar estudios de asociación con enfermedades en todo el genoma, debido a que cumplen la correlación negativa con la distancia. Sin embargo, se observa que la correlación se mantiene por el efecto de los pares fuertemente ligados (véase figura 9, del apartado de *Resultados*), lo que implica que su efecto se extiende a cortas distancias. Así, estos marcadores serían útiles para detectar variantes de enfermedades a distancias cortas. El mantenimiento de la correlación negativa entre el DG entre pares de SNPs y la DF por el efecto de los pares fuertemente ligados ya había sido observada previamente (JEFFREYS *et al.* 2001). Sin embargo, el DG entre pares de microsatélites a lo largo de 11p no sigue el patrón esperado con la distancia por lo que, en principio, estos marcadores no serían predictores adecuados de la posición de los genes de enfermedades en el genoma. Aunque los resultados del estudio de ZAPATA *et al.* (2001b) en 19 cM de la región 11p15 han mostrado que, considerando sólo los alelos más frecuentes, éstos sí podrían predecir la posición de variantes de enfermedades a largas distancias. De esta manera, el clonaje posicional a largas distancias sería posible en una primera fase utilizando como marcadores los alelos más frecuentes de loci microsatélites.

Posteriormente, a partir de una región señalada como candidata en el primer paso, la posición del gen se definiría más con SNPs que marcan regiones de DG más cortas del DNA.

La densidad óptima de los marcadores y el éxito de los estudios de asociación dependen en gran medida de la relación esperada del DG con la DF. Sin embargo, en algunos estudios se ha sugerido que el DG no se correlaciona con la DF en regiones cortas del genoma, por ejemplo de menos de 50 kb (JORDE *et al.* 1994; SCHULZE *et al.* 2002; WALL 2004). Aunque existen muchos otros estudios que no apoyan esta observación y sí obtuvieron correlación a distancias cortas (DUNNING *et al.* 2000; PRITCHARD y PRZEWORSKY 2001). Esta heterogeneidad en los resultados es predecible puesto que a cortas distancias pueden actuar otros fenómenos, además de la recombinación, que pueden afectar al DG, como la mutación o la conversión génica (NAKAJIMA *et al.* 2002; WALL 2004). En el caso de pares de SNPs fuertemente ligados, la conversión génica podría ejercer un efecto similar al de la recombinación, rompiendo las asociaciones. Sin embargo, en nuestros resultados, se observa correlación entre el DG y la FRE o DF entre pares de SNPs, incluso a cortas distancias. Además, el 75 % (3/4) de los pares de SNPs separados por menos de 50 kb, muestran DG significativo con una intensidad promedio, medida en términos de $|D'|$, muy elevada (0,98).

Los niveles de DG parecen diferir mucho a lo largo de los cromosomas, por lo que existe una clara necesidad para la construcción de un mapa de DG en todo el genoma (CLARK *et al.* 1998; HUTTLEY *et al.* 1999; KRUGLYAK 1999; BOEHNKE 2000; TAILLON-MILLER *et al.* 2000; GABRIEL *et al.* 2002; WALL y PRITCHARD 2003). Parte de esta variabilidad se puede explicar por la variación en las tasas de recombinación de varios órdenes de magnitud a lo largo del genoma, desde 0 a 8,8 cM/Mb (Yu *et al.* 2001). Existen evidencias crecientes de que los eventos de recombinación meiótica en los cromosomas humanos tienden a encontrarse agrupados en puntos calientes o *hotspots* de 1 a 2 kb de longitud. Así, los puntos calientes de recombinación tienden a estar separados entre sí por extensos

bloques haplotípicos donde apenas existe recombinación, de manera que los marcadores en esa región presentan un fuerte DG (CHAKRAVARTY *et al.* 1984; TEMPLETON *et al.* 2000; JEFFREYS *et al.* 2000, 2001; YU *et al.* 2001; MAY *et al.* 2002; WALL *et al.* 2003). Dentro de los bloques haplotípicos existe una tendencia de disminución del DG al aumentar la separación entre los marcadores (JEFFREYS *et al.* 2001). Esta diversidad genética estructurada en bloques de DG parece ser una característica común en el genoma humano (DALY *et al.* 2001; GOLDSTEIN 2001; PATIL *et al.* 2001; DAWSON *et al.* 2002; GABRIEL *et al.* 2002; PHILLIPS *et al.* 2003). Así, en los últimos años, ha surgido el proyecto internacional HapMap con el objetivo de caracterizar los patrones de DG entre SNPs a lo largo de todo el genoma, y así facilitar los estudios de asociación para el descubrimiento de genes implicados en enfermedades complejas en humanos (SCHULZE *et al.* 2004; THE INTERNATIONAL HAPMAP CONSORTIUM 2003; ZAPATA *et al.* 2005).

La evaluación del DG de *background* a lo largo del genoma es un punto crítico para una correcta localización, mediante DG, de los genes implicados en enfermedades. Así, la intensidad del DG entre un locus o loci marcadores y un alelo de susceptibilidad a una enfermedad, debe ser mayor que el DG de *background* o DG entre loci marcadores en regiones anónimas del genoma (FREIMER *et al.* 1997; GORDON *et al.* 2000; SERVICE *et al.* 2001). Los resultados obtenidos en el presente trabajo sugieren que los patrones de DG en el brazo corto del cromosoma 11 son dependientes del tipo de marcador utilizado. Por lo tanto, los resultados obtenidos con diferentes tipos de marcadores no son comparables. Este DG dependiente de marcador representa una fracción importante del DG de *background* y parece tener una explicación basada en las diferentes dinámicas mutacionales que poseen los SNPs y los microsatélites. De esta manera, para una mayor comprensión de los patrones de DG a lo largo de los cromosomas humanos será necesario separar el DG dependiente de marcador del DG de *background*. Los SNPs son entidades mutacionales muy pequeñas que aportan poca información por sí mismos. Sin embargo, su baja tasa mutacional supone que se dejan arrastrar por fenómenos evolutivos de fondo que ocurren en el genoma, de manera que son mejores representantes del DG de *background* en

los cromosomas. Los microsatélites por su parte mantienen por cortos espacios de tiempo las señales de la historia evolutiva debido a su elevada tasa de mutación que puede crear asociaciones nuevas y romper otras ya existentes con loci cercanos y lejanos en el cromosoma. Sin embargo, gracias a la diferente antigüedad existente entre los alelos de loci microsatélites, éstos podrían servir para marcar fenómenos en el genoma ocurridos en diferentes épocas de la historia evolutiva.

Se requerirán estudios futuros para confirmar si este patrón de DG dependiente de marcador se encuentra en otros cromosomas o regiones cromosómicas. Además, será necesaria una evaluación en profundidad en otros cromosomas del tipo de marcador genético más idóneo para realizar estudios de asociación con enfermedades complejas en todo el genoma, teniendo en cuenta las dinámicas evolutivas de cada uno.

ESTUDIO DE ASOCIACIÓN DE LOS GENES DEL COLÁGENO TIPO I CON LA OTOSCLEROSIS

En el presente trabajo se realizó un análisis de asociación de la otosclerosis con marcadores polimórficos situados en los genes que codifican para el colágeno tipo I (COL1A1 y COL1A2). Los resultados del presente estudio confirman los resultados previos por parte de MCKENNA *et al.* (1998) que evidenciaron una ausencia de asociación entre el gen COL1A2 y la otosclerosis. Sin embargo, el hallazgo de la asociación del gen COL1A1 con la otosclerosis en ese mismo trabajo, no ha podido ser replicado en nuestra muestra de población gallega. Así, la comparación de las frecuencias alélicas de los marcadores analizados, entre el grupo de afectados por la otosclerosis y el grupo control, no reflejó diferencias significativas en ningún caso ($P \gg 0,05$). Además, la intensidad de asociación entre los alelos de cada marcador con la otosclerosis, medida en términos de coeficientes $|D'|$, reflejó valores muy bajos, que oscilaron entre 0,02 (*RsaI* / COL1A1) y 0,19 (*MspI* / COL1A2), con un valor promedio ($\pm$ SE) de $0,08 \pm 0,02$,

Discusión

muy próximo al valor mínimo que puede tomar $|D'|$. Tampoco reflejaron diferencias significativas las comparaciones entre las frecuencias genotípicas y haplotípicas de pares de loci y la otosclerosis, con valores de probabilidad muy por encima del valor de significación nominal ($P \gg 0,05$).

Estudios previos han sugerido una implicación hormonal en la etiología de la otosclerosis, debido a la observación de una prevalencia dos veces superior en mujeres que en hombres (GORDON 1989; BEN ARAB *et al.* 1993; MENDER y TANGE 2003). Así, para averiguar si el factor sexo estaba implicado en la aparición de la otosclerosis, se realizó el análisis de asociación estratificando la muestra por sexos. Sin embargo, no se evidenciaron diferencias significativas en las frecuencias alélicas de los marcadores entre el grupo de afectados y el grupo control, ni en la muestra de hombres ni en la muestra de mujeres ($P \gg 0,05$).

La falta de potencia estadística podría ser una causa para la obtención de falsos negativos en estudios de asociación cuando, habiendo diferencias genéticas entre los individuos afectados y los individuos control, éstas no son estadísticamente significativas al nivel de significación nominal. En estos casos, permanece una cierta duda de si las diferencias observadas son debidas simplemente a un error de muestreo o si reflejan efectos reales y son no significativas debido a un tamaño de muestra insuficiente. Sin embargo, los resultados evidenciaron sólo pequeñas diferencias en las distribuciones de frecuencias alélicas, genotípicas y haplotípicas de pares de loci, entre los individuos afectados y los individuos control. Estas diferencias, incluso si estuvieran reflejando efectos reales y no fueran debidas a error de muestreo, son seguramente demasiado pequeñas para poseer alguna consecuencia biológica.

Los resultados discrepantes entre diferentes estudios poblacionales de casos-contróles pueden deberse a una variedad de factores entre los que se incluyen, una deficiente selección de los individuos del grupo control, estratificación poblacional, incertidumbres asociadas con la caracterización fenotípica y genotípica, heterogeneidad alélica y de locus, penetrancia reducida,

análisis estadístico inapropiado, interacciones entre factores de riesgo ambientales y/o genéticos, niveles variables de asociaciones no aleatorias (DG) entre los marcadores y las mutaciones subyacentes a la susceptibilidad de la enfermedad, y la historia poblacional (TERWILLIGER y WEISS 1998; COHEN 1999; TODD 1999; JORDE 2000; RISCH 2000; EMAHAZION *et al.* 2001; IOANNIDIS *et al.* 2001; VIELAND 2001; IOANNIDIS *et al.* 2003; LOHMUELLER *et al.* 2003). En el presente trabajo, se trató de minimizar las incertidumbres relacionadas con la caracterización fenotípica y genotípica de los individuos en el estudio de asociación. Específicamente, los criterios diagnósticos para la inclusión de los individuos con otosclerosis fueron estrictos y sólo se aceptaron individuos con otosclerosis confirmada quirúrgicamente que mostraban una hipoacusia de conducción previa característica. Además, la caracterización de los genotipos de los loci marcadores se realizó tratando de minimizar los errores de genotipado. En este sentido, el genotipado de los microsatélites incluidos en el estudio se realizó a través de un sistema electroforético en geles no desnaturalizantes. Este sistema aprovecha el patrón de las bandas minoritarias de los geles en condiciones nativas, como control de un correcto genotipado y minimiza los errores de genotipado de un heterocigoto como homocigoto, debido a una no amplificación total o parcial de uno de los alelos (RODRÍGUEZ *et al.* 2001; RODRÍGUEZ y ZAPATA 2002).

Un requerimiento básico en los estudios poblacionales de casos-contrroles es que los individuos afectados y los individuos control sean seleccionados aleatoriamente de la misma población genéticamente homogénea (CARDON y BELL 2001; FREEDMAN *et al.* 2004). Puesto que, en caso contrario, se podría obtener una asociación espúrea por la existencia de unas diferencias en las frecuencias alélicas entre casos y controles, debidas únicamente a una estratificación o mezcla poblacional subyacente más que a la proximidad de la mutación de la enfermedad. En el presente estudio las muestras de individuos afectados por otosclerosis y de individuos control fueron correctamente seleccionadas de la misma población. Además, se realizó un análisis de la estructura genética poblacional en las muestras de enfermos, controles y muestra total. En ningún caso se observaron desviaciones significativas de las frecuencias genotípicas

Discusión

observadas respecto a las proporciones Hardy-Weinberg, lo que parece indicar que estamos ante una población genéticamente homogénea sin una estructuración subyacente aparente.

McKENNA *et al.* (1998) investigaron las bases genéticas de la otosclerosis realizando un estudio de asociación, utilizando para ello tres marcadores polimórficos localizados en el gen COL1A1, con individuos afectados e individuos control, aleatorios. Sus resultados sugerían la existencia de asociación entre el gen COL1A1 y la otosclerosis. Así, estas observaciones se contraponen a nuestros resultados. Sin embargo, la evaluación del diseño del estudio llevado a cabo por McKENNA *et al.* (1998), sugiere que los sujetos control podrían no ser los adecuados para realizar el estudio de asociación. En el estudio realizado por estos autores, las frecuencias alélicas de pacientes con otosclerosis fueron obtenidas de la población de Massachussets (USA) (Massachussets Eye and Ear Infirmary, MEEI, <http://www.meei.harvard.edu/research/docs/hscappl.doc>). Estas frecuencias alélicas se compararon con frecuencias alélicas de controles del *Centre d'Étude du Polymorphisme Humain* (CEPH, <http://www.cephb.fr/cephdb>) y con datos europeos reflejados en *Genome Database* (GDB, <http://www.gdb.org>). Estos grupos control provienen de poblaciones diferentes a la del grupo de afectados. Anteriormente se señaló la importancia de la elección del grupo control en estudios de asociación que, en caso contrario, podría causar un error de tipo I (rechazar la hipótesis nula de no asociación siendo verdadera). Así, en nuestro estudio el grupo control proviene de la misma población genéticamente homogénea del grupo de afectados. Con el fin de investigar cómo la utilización de diferentes grupos control afectaría las conclusiones obtenidas por McKENNA *et al.* (1998), se realizó un análisis de asociación a partir de las frecuencias alélicas de los individuos afectados reflejadas en el trabajo de McKENNA *et al.* (1998) y nuestro propio grupo control de la población gallega. La tabla 57 muestra el resultado de este análisis de asociación con nuestro grupo control, así como la comparación de los resultados obtenidos con los controles de CEPH y GDB. Se observa que las asociaciones significativas obtenidas por McKENNA *et al.* (1998) (utilizando individuos de CEPH y GDB como controles para la asociación con los

marcadores *RsaI* y *MnII*, respectivamente) disminuyeron en intensidad y resultaron no significativas al utilizar nuestra muestra control de población gallega, igualmente incorrecta. De esta manera, los resultados del trabajo de McKENNA *et al.* (1998) varían utilizando diferentes muestras control, igualmente incorrectas, lo que sugiere que el resultado de asociación obtenido por estos autores está causado por un error de tipo I.

Tabla 57. Análisis de asociación entre los marcadores *RsaI* y *MnII* (en *COL1A1*) y la Otosclerosis, con los datos de individuos afectados de McKENNA *et al.* (1998) y a) Controles utilizados por McKENNA *et al.* (1998) o b) Controles de la población gallega.

Marcadores	Muestra control					
	CEPH / GDB			Galicia		
	D' ($\pm$ SE)	χ^2 (g.l. = 1)	P	D' ($\pm$ SE)	χ^2 (g.l. = 1)	P
<i>RsaI</i>	-0,342 $\pm$ 0,140	5,01	0,025	-0,070 $\pm$ 0,081	0,75	0,386
<i>MnII</i>	0,181 $\pm$ 0,070	6,80	0,009	0,116 $\pm$ 0,066	3,07	0,080

Nuestro trabajo no presenta evidencias de una implicación de los genes del colágeno tipo I en la etiología de la otosclerosis. Sin embargo, no se puede descartar la existencia de una implicación del gen *COL1A1* en la etiología de la otosclerosis, en la población estudiada por McKENNA *et al.* (1998). Diversos factores, como la naturaleza genéticamente heterogénea de la otosclerosis, influencias medioambientales, o variaciones en el grado del DG entre los marcadores y las mutaciones de la enfermedad a través de las poblaciones, podrían explicar las discrepancias entre diferentes estudios. Así mismo, es igualmente importante señalar que tanto el presente estudio como el de McKENNA *et al.* (1998) no detectaron ninguna asociación de los polimorfismos del gen *COL1A2* con la otosclerosis. Sin embargo, este hecho no excluye definitivamente al gen *COL1A2* como un gen candidato. Serán necesarios estudios posteriores en

diferentes poblaciones para averiguar si los genes que codifican para el colágeno tipo I, juegan o no algún papel en la etiología de la otosclerosis.

La naturaleza heterogénea de la otosclerosis se refleja claramente en los cuatro loci de susceptibilidad localizados hasta el momento en los cromosomas 15 (OTSC1, Tomek *et al.* 1998), 7 (OTSC2, Van Den Bogaert *et al.* 2001), 6 (OTSC3, Chen *et al.* 2002; Van Den Bogaert *et al.* 2002) y 3 (OTSC5, Van Den Bogaert *et al.* 2004). Cada uno de ellos ha sido localizado mediante análisis de ligamiento en familias de procedencia india, belga, chipriota y holandesa, respectivamente. Además, previamente al hallazgo de OTSC5, la *Human Genome Organisation* (HUGO) informó de la existencia de un cuarto locus de susceptibilidad a la otosclerosis (OTSC4). Sin embargo, la localización de este locus en el genoma todavía no se ha publicado. Dada la heterogeneidad genética que parece mostrar la otosclerosis, cabría la posibilidad de que estos loci de susceptibilidad no estuvieran implicados en la aparición de la enfermedad en la población gallega. Los estudios de asociación para la localización de genes implicados en enfermedades han ganado popularidad sobre los análisis de ligamiento convencionales, puesto que, en principio, ofrecen mayor potencia para detectar la presencia de genes con un efecto modesto en enfermedades complejas (Risch y Merikangas 1996; Ardlie *et al.* 2002). Así, sería interesante ensayar en la población gallega la implicación de los loci de susceptibilidad a la otosclerosis, OTSC1, 2, 3 y 5; mediante análisis de asociación.

6. Conclusiones

I. Evaluación del DG dependiente de marcador en el brazo corto del cromosoma 11 humano

1. La frecuencia y magnitud del DG de *background* a lo largo de 11p resultó diferente según el marcador genético considerado. El DG resultó más frecuente entre pares de microsatélites que entre pares de SNPs. Sin embargo, en general, los SNPs se asociaron entre sí con mayor intensidad que los microsatélites. Los resultados sugieren que la diferente dinámica mutacional de cada marcador es el factor determinante de las diferencias observadas en los patrones de DG entre SNPs y entre microsatélites.
2. Los DG interalélicos entre pares de microsatélites a lo largo de 11p siguen un patrón dependiente del tamaño y frecuencia de los alelos implicados en cada haplotipo. Así, se confirma el patrón de DG dependiente de alelo observado previamente en la región cromosómica 11p15 y en Xq25-28.
3. Las distribuciones del DG entre pares de SNPs y entre pares de microsatélites siguieron patrones bien diferenciados a lo largo de 11p. La mayoría de los pares de SNPs en DG estuvieron separados por menos de 0,5 cM. Por el contrario, el DG entre pares de microsatélites se extiende por toda la región analizada con FRE máximas próximas a la segregación independiente (42 cM).
4. La intensidad del DG entre los pares de SNPs sigue una correlación negativa significativa con la FRE. Por el contrario, la intensidad del DG entre pares de microsatélites no parece seguir ninguna tendencia clara al aumentar la FRE. Aunque al considerar sólo los alelos más frecuentes de los microsatélites, sí existe una tendencia hacia una disminución de la intensidad del DG a medida que aumenta la FRE.
5. Los resultados sugieren una posible estrategia de clonación posicional de genes de enfermedades en dos fases. Una primera fase podría realizarse

Conclusiones

con loci microsatélites cuyos DG abarcan extensas regiones en el genoma. Pero este primer paso se haría sólo con los alelos más frecuentes, ya que parecen predecir mejor la posición de los genes en los cromosomas. En una segunda fase, la posición del gen se delimitaría con mayor precisión mediante SNPs, caracterizados por presentar DG a menores distancias cromosómicas.

6. Los patrones de DG de *background* en 11p son dependientes del tipo de marcador genético considerado. Los resultados sugieren que los SNPs son mejores indicadores que los microsatélites del DG de *background* en los cromosomas, al no estar contaminados por una elevada tasa de mutación. Se requerirán estudios futuros para averiguar si este patrón de DG dependiente de marcador se confirma en otros cromosomas.

II. Estudio de asociación de la otosclerosis con los genes del colágeno tipo I, COL1A1 y COL1A2

1. La comparación de las frecuencias alélicas, genéticas y haplotípicas de pares de loci entre el grupo de afectados y el grupo control no reflejó diferencias significativas en ningún caso. Así mismo, al estratificar la muestra por sexos tampoco se detectaron diferencias significativas de las frecuencias alélicas entre afectados y controles, ni en la muestra de mujeres, ni en la muestra de hombres.
2. No se detectó asociación entre la otosclerosis y los polimorfismos analizados en los genes COL1A1 y COL1A2 en la muestra de población gallega. Por lo tanto, este trabajo no permite confirmar la asociación previamente detectada entre la otosclerosis y el gen COL1A1 en una población de Massachussets (USA).
3. La ausencia de asociación no parece ser un problema de baja potencia estadística. Sin embargo, la evaluación del estudio de asociación realizado

en la población de Massachussets (USA) revela la utilización de una población control inadecuada.

4. No obstante, no se puede descartar la implicación de los genes del colágeno tipo I en la etiología de la otosclerosis, dado que puede existir heterogeneidad genética tanto dentro, como entre poblaciones. En el futuro, se requerirán más estudios que esclarezcan si los genes del colágeno tipo I están asociados con la otosclerosis.

7. Bibliografía

- ABECASIS G. R., NOGUCHI E., HEINZMANN A., TRAHERNE J. A., BHATTACHARYYA S., LEAVES N. I., ANDERSON G. G., ZHANG Y., LENCH N. J., CAREY A., CARDON L. R., MOFFATT M. F., COOKSON W. O. 2001. Extent and distribution of linkage disequilibrium in three genomic regions. *Am. J. Hum. Genet.* **68** (1): 191 – 197.
- AGRESTI A. 1990. *Categorical Data Analysis*. John Wiley & Sons, New York.
- AKAI J., KIMURA A. & HATA R-I. 1999. Transcriptional regulation of the human type I collagen $\alpha 2$ (COL1A2) gene by the combination of two dinucleotide repeats. *Gene* **239**: 65 – 73.
- AKEY J. M., ZHANG G., ZHANG K. JIN L. & SHRIVER M. D. 2002. Interrogating a high-density SNP map for signatures of natural selection. *Genome Res.* **12**: 1805 – 1814.
- ALLARD R. W., JAIN S. K. & WORKMAN P. L. 1968. The genetics of inbreeding populations. *Adv. Genet.* **14**: 55 – 131.
- ALTMAN F., GLASGOLD A. & MACDUFF J. P. 1967. The incidence of otosclerosis as related to race and sex. *Ann. Otol. Rhinol. Laryngol.* **76**: 377 – 392.
- ALTMÜLLER J., PALMER L. J., FISCHER G., SCHERB H., WJST M. 2001. Genomewide scans of complex human diseases: true linkage is hard to find. *Am. J. Hum. Genet.* **69** (5): 936 – 50.
- ANDERSON T. J. C., PAUL R. E. L., DONELLY C. A., DAY K. P. 2000. Do malaria parasites mate non-randomly in the mosquito midgut?. *Genet. Res.* **75**: 285 – 296.
- ANDERSON E. C. & SLATKIN M. 2004. Population-genetic basis of haplotype blocks in the 5q31 region. *Am. J. Hum. Genet.* **74**: 40 – 49.
- ARDLIE K. G., KRUGLYAK L., SEIELSTAD M. 2002. Patterns of linkage disequilibrium in the human genome. *Nature Reviews Genetics* **3**: 299 – 309.
- ARNOLD W., NIEDERMEYER H. P., LEHN N., NEUBERT W., HÖFLER H. 1996. Measles virus in otosclerosis and the specific immune response of the inner ear. *Acta Otolaryngol. (Stockh)*. **116**: 705 – 709.
- ASMUSSEN M. A. y CLEGG M. T. 1981. Dynamics of the linkage disequilibrium function under models of gene-frequency hitchhiking. *Genetics* **99**: 337 – 356.
- AULCHENKO Y. S., HEUTINK P., MACKAY I., BERTOLI-AVELLA A. M., PULLEN J., VAESSEN N., RADEMAKER T. A., SANDKUIJL L. A., CARDON L., OOSTRA B., VAN DUIJN C. M. 2004. Linkage disequilibrium in young genetically isolated Dutch population. *Eur. J. Hum. Genet.* **12** (7): 527 – 34.
- AVERY P. J. & HILL W. G. 1979. Distribution of linkage disequilibrium with selection and finite population size. *Genet. Res.* **33**: 29 – 48.
- BAKER R., LYNCH J., FERGUSON L., PRIESTLEY L., SYKES B. 1991. PCR detection of five restriction site dimorphisms at the type I collagen loci COL1A1 and COL1A2. *Nucleic Acids Res.* **19** (15): 4315.

Bibliografia

- BARKER J. S. F. 1979. Inter-locus interactions: A review of experimental evidence. *Theor. Pop. Biol.* **16**: 323 – 346.
- BELL G. I. & JURKA J. 1997. The length distribution of perfect dimer repetitive DNA is consistent with its evolution by an unbiased single-step mutation process. *J. Mol. Evol.* **44**(4): 414 – 21.
- BEN ARAB S. BONAÏTI-PELLIE C., BELKAHIA A. 1993. A genetic study of otosclerosis in a population living in the north of Tunisia. *Ann. Genet.* **36**: 111 – 116.
- BENNETT S. T., LUCASSEN A. M., GOUGH S. C., POWELL E. E., UNDLIEN D. E., PRITCHARD L. E., MERRIMAN M. E., KAWAGUCHI Y., DRONSFIELD M. J., POCIOT F., NERUP J., BOUZEKRI N., CAMBON-THOMSEN A., RONNINGEN K. S., BARNETT A. H., BAIN S. C. & TODD J. A. 1995. Susceptibility to human type 1 diabetes at IDDM2 is determined by tandem repeat variation at the insulin gene minisatellite locus. *Nat. Genet.* **9** (3): 284 – 92.
- BERNARDI G. 1995. The human genome: Organization and evolutionary history. *Annu. Rev. Genetics* **29**: 445 – 476.
- BERNSTEIN J. M., SHANAHAN T. C., SCHAFER F. M. 1996. Further observations on the role of the MHC genes and certain hearing disorders. *Acta Otolaryngol. (Stockh)*. **116**: 666 – 671.
- BODMER W. F. & PARSONS P. A. 1962. Linkage and recombination in evolution. *Advan. Genet.* **11**: 1 – 100.
- BODMER W. F. 1986. Human Genetics: the Molecular Challenge. *Cold. Spring. Harb. Symp. Quant. Biol.* **51**: 1 – 13.
- BOEHNKE M. 2000. A look at linkage disequilibrium. *Nat. Genet.* **25** :246 – 247.
- BONNEN P. E., STORY M. D., ASHORN C. L., BUCHHOLZ T. A., WEIL M. M., NELSON D. L. 2000. Haplotypes at ATM identify coding-sequence variation and indicate a region of extensive linkage disequilibrium. *Am. J. Hum. Genet.* **67** (6): 1437 – 51.
- BOTSTEIN D. y RISCH N. 2003. Discovering genotypes underlying human phenotypes: Past successes for mendelian disease, future approaches for complex disease. *Nature Genetics* **33**: 228 – 237.
- BOTSTEIN D., WHITE R. L., SKOLNICK M. & DAVIS R. W. 1980. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *Am. J. Hum. Genet.* **32**: 314 – 331.
- BRINKMANN B., KLINTSCHAR M., NEUHUBER F., HUHNE J., ROLF B. 1998. Mutation rate in human microsatellites: influence of the structure and length of the tandem repeat. *Am. J. Hum. Genet.* **62** (6): 1408 – 15.
- BROMAN KW, WEBER JL. 2000. Characterization of human crossover interference. *Am. J. Hum. Genet.* **66**: 1911 – 26.
- BROOKES A. J. 1999. The essence of SNPs. *Gene* **234**: 177 – 186.

- BROSNAN M., BURNS H., JAHN A. F., HAWKE M. 1977. Surgery and histopathology of the stapes in osteogenesis imperfecta tarda. *Arch. Otolaryngol.* **103**: 294 – 298.
- BROWN A. H. D. 1975. Sample sizes required to detect linkage disequilibrium between two or three loci. *Theor. Pop. Biol.* **8**: 184 – 201.
- BUETOW K. H., SHIANG R., YANG P., NAKAMURA Y., LATHROP M., WHITE R., WASMUTH J. J., WOOD S., BERDAHL L. D., LEYSENS N. J., RITTY T. M., WISE M. E. & MURRAY C. 1991. A detailed multipoint map of human chromosome 4 provides evidence for linkage heterogeneity and position-specific recombination rates. *Am. J. Hum. Genet.* **48**: 911 – 925.
- BYRNE J. A. & SMITH P. J. 1993. The 11p15.5 ribonucleotide reductase M1 subunit locus is not imprinted in Wilm's tumour and hepatoblastoma. *Hum. Genet.* **91**: 275 – 277.
- CALABRESE P. P., DURRETT R. T., AQUADRO C. F. 2001. Dynamics of microsatellite divergence under stepwise mutation and proportional slippage/point mutation models. *Genetics* **159** (2): 839 – 52.
- CALABRESE P., DURRETT R. 2003. Dinucleotide repeats in the Drosophila and human genomes have complex, length-dependent mutation processes. *Mol. Biol. Evol.* **20** (5): 715 – 25.
- CAMP N. J., SWENSEN J., HORNE B. D., FARNHAM J. M., THOMAS A., CANNON-ALBRIGHT L. A. & TAVTIGIAN S. V. 2005. Characterization of linkage disequilibrium structure, mutation history, and tagging SNPs, and their use in association analyses: *ELAC2* and familial early-onset prostate cancer. *Genet. Epidemiol.* **28**: 232 – 243.
- CARDON L. R., BELL J. I. 2001. Association study designs for complex diseases. *Nat. Rev. Genet.* **2** (2): 91 – 9.
- CARGILL M., ALTSHULER D., IRELAND J., SKLAR P., ARDLIE K., PATIL N., SHAW N., LANE C. R., LIM E. P., KALYANARAMAN N., NEMESH J., ZIAUGRA L., FRIEDLAND L., ROLFE A., WARRINGTON J., LIPSHUTZ R., DALEY G. Q., LANDER E. S. 1999. Characterization of single-nucleotide polymorphisms in coding regions of human genes. *Nat. Genet.* **22** (3): 231 – 8.
- CARLSON C. S., EBERLE M. A., RIEDER M. J., YI Q., KRUGLYAK L., NICKERSON D. A. 2004. Selecting a maximally informative set of single-nucleotide polymorphisms for association analyses using linkage disequilibrium. *Am. J. Hum. Genet.* **74** (1): 106 – 20.
- CAVALLI-SFORZA L., MENOZZI P. & PIAZZA A. 1994. *The History and Geography of Human Genes*. Princeton University Press. New Jersey. USA.
- CHAKRABORTY R. 1984. Detection of nonrandom association of alleles from the distribution of the number of heterozygous loci in a sample. *Genetics* **108**: 719 – 731.
- CHAKRABORTY R., KIMMEL M., STIVERS D. N., DAVISON L. J., DEKA R. 1997. Relative mutation rates at di-, tri-, and tetranucleotide microsatellite loci. *Proc. Natl. Acad. Sci. U S A.* **94** (3): 1041 – 6.

Bibliografía

- CHAKRABORTY R. & WEISS K. M. 1988. Admixture as a tool for finding linked genes and detecting that difference from allelic association between loci. *Proc. Natl. Acad. Sci. U S A.* **85** (23): 9119 – 23.
- CHAKRABORTY R. & DANKER-HOPFE H. 1991. Analysis of Population Structure: A Comparative Study of Different Estimations of Wright's Fixation Indices, pp203-254. En: Handbook of Statistics. Vol 8. Editado por Rao, C. R. y Chakraborty, R. Elsevier.
- CHAKRAVARTY A. 1998. It's raining SNPs, hallelujah?. *Nature Genetics* **19**: 216 – 217.
- CHAKRAVARTY A. 1999. Population genetics – making sense out of sequence. *Nature Genetics* **21**: 56 – 60.
- CHAKRAVARTY A., BUETOW K. H., ANTONARAKIS S. E., WABER P. G., BOEHM C. D. & KAZAZIAN H. H. 1984. Nonuniform recombination within the human β -globin gene cluster. *Am. J. Hum. Genet.* **36**: 1239 – 1258.
- CHAPMAN N. H. & WIJSMAN E. M. 1998. Genome screens using linkage disequilibrium tests: Optimal marker characteristics and feasibility. *Am. J. Hum. Genet.* **63**: 1872 – 1885.
- CHARLESWORTH B., MORAN M. y CHARLESWORTH D. 1993. The effect of deleterious mutations on neutral molecular variation. *Genetics* **134**: 1289 – 1303.
- CHARLESWORTH D., CHARLESWORTH B., MORGAN M. T. 1995. The pattern of neutral molecular variation under the background selection model. *Genetics* **141** (4): 1619 – 32.
- CHEN W., CAMPBELL C. A., GREEN G. E., VAN DEN BOGAERT K., KOMODIKIS C., MANOLIDIS L. S., ACONOMOU E., KYAMIDES Y., CHRISTODOULOU K., FAGHEL C., GIGUÈRE C. M., ALFORD R. L., MANOLIDIS S., VAN CAMP G., SMITH R. J. H. 2002. Linkage of otosclerosis to a third locus (OTSC3) on human chromosome 6p21.3-22.3. *J. Med. Genet.* **39**: 473 – 477.
- CHEN W., MCKENNA M. J., VAN CAMP G., SMITH R. J. H. 2002. Otosclerosis and COL1A1. (Abstract). The American Society of Human Genetics, 52nd Annual Meeting, Baltimore.
- CHI D. D., HING A. V., HELMS C., STEINBRUECK T., MISHRA S. K., DONIS-KELLER H. 1992. Two chromosome 7 dinucleotide repeat polymorphisms at gene loci epidermal growth factor receptor (EGFR) and pro alpha 2 (I) collagen (COL1A2). *Hum. Mol. Genet.* **1** (2): 135.
- CLARK A. G., WEISS K. M., NICKERSON D. A., TAYLOR S. L., BUCHANAN A., STENGARD J., SALOMAA V., VARTIAINEN E., PEROLA M., BOERWINKLE E., SING C. F. 1998. Haplotype structure and population genetic inferences from nucleotide-sequence variation in human lipoprotein lipase. *Am. J. Hum. Genet.* **63** (2): 595 – 612.
- CLARK A. G., NIELSEN R., SIGNOROVITCH J., MATISE T. C., GLANOWSKI S., HEIL J., WINN-DEEN E. S., HOLDEN A. L., LAI E. 2003. Linkage disequilibrium and inference of ancestral recombination in 538 single-nucleotide polymorphism clusters across the human genome. *Am. J. Hum. Genet.* **73** (2): 285 – 300.
- CLEGG M. T. 1984. Dynamics of multilocus genetic systems, en Oxford Surveys in Evolutionary Biology, vol. 1, pp. 160 – 183. Ed. R. Dawkins y M. Rittley. Oxford University Press. Oxford, U.K.

- COHEN B. 1999. Freely associating. *Nat. Genet.* **22**: 1 – 2.
- COLLINS F. S. 1995. Positional cloning moves from perditiional to traditional. *Nature Genetics* **9** (4): 347 – 50.
- COLLINS A., FREZAL J., TEAGUE J., MORTON N. E. 1996. A metric map of humans: 23,500 loci in 850 bands. *Proc. Natl. Acad. Sci. U S A* **93** (25): 14771 – 14775.
- COLLINS A., LONJOU C. & MORTON N. E. 1999. Genetic epidemiology of single-nucleotide polymorphisms. *Proc. Natl. Acad. Sci. USA* **96**: 15173 – 15177.
- CONSTANTINOU C. D., SPOTILA L. D., ZHUANG J., SEREDA L., HANNING C., PROCKOP D. J. 1990. PvuII polymorphism at the COL1A2 locus. *Nucleic Acids Res.* **18** (18): 5577.
- CROUAEU-ROY B., SERVICE S., SLATKIN M. & FREIMER N. 1996. A fine-scale comparison of the human and chimpanzee genomes: linkage, linkage disequilibrium and sequence analysis. *Hum. Mol. Genet.* **5**: 1131 – 1137.
- CUI X. F., LI H. H., GORADIA T. M., LANGE K., KAZAZIAN H. H. JR, GALAS D., ARNHEIM N. 1989. Single-sperm typing: determination of genetic distance between the G gamma-globin and parathyroid hormone loci by using the polymerase chain reaction and allele-specific oligomers. *Proc. Natl. Acad. Sci. U S A.* **86** (23): 9389 – 93.
- CURIE-COHEN M. 1982. Estimates of Inbreeding in a Natural Population: A Comparison of Sampling Properties. *Genetics* **100**: 339 – 358.
- D'AGOSTINO R. B., CHASE W. & BELANGER A. 1988. The appropriateness of some common procedures for testing the equality of two independent binomial populations. *The American Statistician* **42**: 198 – 202.
- DALY M. J., RIOUX J. D., SCHAFFNER S. F., HUDSON T. J. & LANDER E. S. 2001. High-resolution haplotype structure in the human genome. *Nat. Genet.* **29**: 229 – 232.
- DAWSON E., ABECASIS G. R., BUMPSTEAD S., CHEN Y., HUNT S., BEARE D. M., PABIAL J., DIBLING T., TINSLEY E., KIRBY S., CARTER D., PAPASPYRIDONOS M., LIVINGSTONE S., GANSKE R., LOHMUSSEAR E., ZERNANT J., TONISSON N., REMM M., MAGI R., PUURAND T., VILO J., KURG A., RICE K., DELOUKAS P., MOTT R., METSPALU A., BENTLEY D. R., CARDON L. R., DUNHAM I. 2002. A first generation linkage disequilibrium map of human chromosome 22. *Nature* **418**: 544 – 548.
- DEKA R., JIN L., SHRIVER M. D., YU L. M., DECROO S., HUNDRIESER J., BUNKER C. H., FERRELL R. E., CHAKRABORTY R. 1995. Population genetics of dinucleotide (dC-dA)_n(dG-dT)_n polymorphisms in world populations. *Am. J. Hum. Genet.* **56** (2): 461 – 74.
- DEMPSTER A. P., LAIRD N. M. & RUBIN D. B. 1977. Maximum likelihood from incomplete data via de EM algorithm. *J. Roy. Statist. Soc.* **39**: 1 – 38.
- DE LA CHAPELLE A., WRIGHT F. A. 1998. Linkage disequilibrium mapping in isolated populations: the example of Finland revisited. *Proc. Natl. Acad. Sci. U S A* **95** (21): 12416 – 12423.

Bibliografia

- DEVLIN B. & RISCH N. 1995. A comparison of linkage disequilibrium measures for fine-scale mapping. *Genomics* **29**: 311 – 322.
- DEVLIN B., RISCH N. & ROEDER K. 1996. Disequilibrium mapping: Composite likelihood for pairwise disequilibrium. *Genomics* **36**: 1 – 16.
- DIB C., FAURE S., FIZAMES C., SAMSON D., DROUOT N., VIGNAL A., MILLASSEAU P., MARC S., HAZAN J., SEBOUN E., LATHROP M., GYAPAY G., MORISSETTE J., WEISSENBAACH J. 1996. A comprehensive genetic map of the human genome based on 5,264 microsatellites. *Nature* **380** (6570): 152 – 4.
- DIETRICH W. F., MILLER J. C., STEEN R. G., MERCHANT M., DAMRON-BOLES D., HUSAIN Z., DREDGE, R., DALY M. J., INGALLS K. A., O'CONNOR T. J., EVANS C. A., DEANGELIS M. M., LEVINSON D. M., KRUGLYAK L., GODDMAN N., COPELAND N. G., JENKINS N. A., HAWKINS T. L., STEIN L., PAGE D. C. & LANDER E. S. 1996. A Comprehensive Genetic Map of the Mouse Genome. *Nature* **380**: 149 – 152.
- DI RIENZO A., PETERSON A. C., GARZA J. C., VALDES A. M., SLATKIN M., FREIMER N. B. 1994. Mutational processes of simple-sequence repeat loci in human populations. *Proc. Natl. Acad. Sci. U S A* **91** (8): 3166 – 70.
- DI RIENZO A., DONNELLY P., TOOMAJIAN C., SISK B., HILL A., PETZL-ERLER M. L., HAINES G. K., BARCH D. H. 1998. Heterogeneity of microsatellite mutations within and between loci, and implications for human demographic histories. *Genetics* **148** (3): 1269 – 84.
- DUNNING A. M., DUROCHER F., HEALEY C. S., TEARE M. D., MCBRIDE S. E., CARLOMAGNO F., XU C. F., DAWSON E., RHODES S., UEDA S., LAI E., LUBEN R. N., VAN RENSBURG E. J., MANNERMAA A., KATAJA V., RENNART G., DUNHAM I., PURVIS I., EASTON D., PONDER B. A. 2000. The extent of linkage disequilibrium in four populations with distinct demographic histories. *Am. J. Hum. Genet.* **67** (6): 1544 – 54.
- EAVES A. I., MERRIMAN T. R., BARBER R. A., NUTLAND S., TUOMILEHTO-WOLF J. T., CUCA F. & TODD J. A. 2000. The Genetically Isolated Populations of Finland and Sardinia May not be a Panacea for Linkage Disequilibrium Mapping of Common Disease Genes. *Nature Genetics* **25**: 320 – 323.
- EISENBARTH I., VOGEL G., KRONE W., VOGEL W., ASSUM G. 2000. An isochore transition in the NF1 gene region coincides with a switch in the extent of linkage disequilibrium. *Am. J. Hum. Genet.* **67** (4): 873 – 80.
- EISENBARTH I., STRIEBEL A. M., MOSCHGATH E., VOGEL W., ASSUM G. 2001. Long-range sequence composition mirrors linkage disequilibrium pattern in a 1.13 Mb region of human chromosome 22. *Hum. Mol. Genet.* **10** (24): 2833 – 9.
- ELLEGREN H. 2004. Microsatellites: simple sequences with complex evolution. *Nature Rev. Genet.* **5**: 435 – 445.
- ELSTON R. C. & FORTHOFFER R. 1977. Testing for Hardy-Weinberg equilibrium in small samples. *Biometrics* **33**: 536 – 542.
- EMAHAZION T., FEUK L., JOBS M., SAWYER S. L., FREDMAN D., ST CLAIR D., PRINCE J. A., BROOKES A. J. 2001. SNP association studies in Alzheimer's disease highlight problems for complex disease analysis. *Trends Genet.* **17** (7): 407 – 13.

- EMIGH T. H. 1980. A Comparison of tests for Hardy-Weinberg equilibrium. *Biometrics* **36**: 627 – 642.
- EMIGH T. H. & KEMPTHORNE O. 1975. A note on goodness of fit of a population to Hardy-Weinberg structure. *Am. J. Hum. Genet.* **27**: 778 – 783.
- EVERITT B. S. 1997. *The Analysis of Contingency Tables*. Chapman & Hall, London.
- EXCOFFIER L. & SLATKIN M. 1995. Maximum-likelihood estimation of molecular haplotype frequencies in a diploid population. *Mol. Biol. Evol.* **12** (5): 921 – 927.
- FALLIN D., COHEN A., COHEN A., ESSIUX L., CHUMAKOV I., BLUMENFELD M., COHEN D., SCHORK N. J. 2001. Genetic analysis of case/control data using estimated haplotype frequencies: Application to APOE locus variation and Alzheimer's disease. *Genome Res.* **11**: 143 – 151.
- FALLIN D. & SCHORK N. J. 2000. Accuracy of haplotype frequency estimation for biallelic loci, via the expectation-maximization algorithm for unphased diploid genotype data. *Am J Hum. Genet.* **67** (4): 947 – 59.
- FARRALL M. & WEEKS D. E. 1998. Mutational mechanisms for generating microsatellite allele-frequency distributions: An analysis of 4.558 markers. *Am. J. Hum. Genet.* **62**: 1260 – 1262.
- FELDMAN M. W., BERGMAN A., POLLOCK D. D., GOLDSTEIN D. B. 1997. Microsatellite genetic distances with range constraints: analytic description and problems of estimation. *Genetics* **145** (1): 207 – 16.
- FELSENSTEIN, J. 1979. A Mathematically Tractable Family of Genetic Mapping Functions with Different Amounts of Interference. *Genetics* **91**: 769 – 775.
- FORTON J., KWIATKOWSKI D., ROCKETT K., LUONI G., KIMBER M., HULL J. 2005. Accuracy of haplotype reconstruction from haplotype-tagging single-nucleotide polymorphisms. *Am. J. Hum. Genet.* **76**(3): 438 – 48.
- FOX D. P. 1973. The control of chiasma distribution in the locus *Schistocera gregaria* (Forsk.). *Chromosoma* **43**: 289 – 328.
- FRANKLIN I. y LEWONTIN R. C. 1970. Is the gene the unit of selection?. *Genetics* **65**: 707 – 734.
- FREEDMAN M. L., REICH D., PENNEY K. L., McDONALD G. J., MIGNAULT A. A., PATTERSON N., GABRIEL S. B., TOPOL E. J., SMOLLER J. W., PATO C. N., PATO M. T., PETRYSHEN T. L., KOLONEL L. N., LANDER E. S., SKLAR P., HENDERSON B., HIRSCHHORN J. N., ALTSCHULER D. 2004. Assessing the impact of population stratification on genetic association studies. *Nat. Genet.* **36** (4): 388 – 93.
- FREIMER N. B., SERVICE S. K., SLATKIN M. 1997. Expanding on population studies. *Nature Genetics* **17** (4): 371 – 3.

Bibliografía

- FRIEDMAN T. B., LIANG Y., WEBER J. L., HINNANT J. T., BARBER T. D., WINATA S., NYOMAN ARHYA I. y ASHER J. H. 1995. A gene for congenital, recessive deafness DFNB3 maps to the pericentromeric region of chromosome 17. *Nature Genetics* **9**: 86 – 91.
- FU Y. X. & ARNOLD J. 1992. A Table of exact sample sizes for use with Fisher's Exact Test for 2 x 2 tables. *Biometrics* **48**: 1103 – 1112.
- GABRIEL S. B., SCHAFFNER S. F., NGUYEN H., MOORE J. M., ROY J., BLUMENSTIEL B., HIGGINS J., DEFELICE M., LOCHNER A., FAGGART M., LIU-CORDERO S. N., ROTIMI C., ADEYEMO A., COOPER R., WARD R., LANDER E. S., DALY M. J. & ALTSHULER D. 2002. The structure of haplotype blocks in the human genome. *Science* **296**: 2225 – 2229.
- GARZA J. C., SLATKIN M., FREIMER N. B. 1995. Microsatellite allele frequencies in humans and chimpanzees, with implications for constraints on allele size. *Mol. Biol. Evol.* **12** (4): 594 – 603.
- GODDARD K. A. B., HOPKINS P. J., HALL J. M. & WITTE J. S. 2000. Linkage Disequilibrium and Allele-Frequency Distributions for 114 Single-Nucleotide Polymorphisms in Five Populations. *Am. J. Hum. Genet.*, **66**: 216 – 234.
- GOLDSTEIN D. B. 2001. Islands of linkage disequilibrium. *Nat. Genet.* **29**: 109 – 111.
- GOLDSTEIN D. B. 2003. Pharmacogenetics in the laboratory and the clinic. *N. Engl. J. Med.* **348**: 553 – 556.
- GOLDSTEIN D. B., AHMADI K. R., WEALE M. E. & WOOD N. W. 2003. Genome scans and candidate gene approaches in the study of common diseases and variable drug responses. *Trends in Genetics* **19**: 615 – 622.
- GORDON D., SIMONIC I. & OTT J. 2000. Significant Evidence for Linkage Disequilibrium over a 5-cM Region Among Afrikaners. *Genetics* **66**: 8 7- 92.
- GORDON M. A. 1989. The genetics of otosclerosis: a review. *Am. J. Otol.* **10**: 426 – 438.
- GORDON M. A., MCPHEE J. R., VAN DE WATER T. R., RUBEN R. J. 1992. Aberration of the tissue collagenase system in association with otosclerosis. *Am. J. Otol.* **13**: 398 – 407.
- GRAY I. C., CAMPBELL D. A. & SPURR N. K. 2000. Single nucleotide polymorphisms as tools in human genetics. *Hum. Mol. Genet.* **9**: 2403 – 2408.
- GRAYELI A. B., ESCOUBET B., BICHARA M., NICOLAS J., SILVE C. FRIEDLANDER G., STERKERS O., FERRARY E. 2003. Increased activity of the diastrophic dysplasia sulphate transporter in otosclerosis and its inhibition by sodium fluoride. *Otol. & Neurotol.* **24**: 854 – 862.
- GRAYELI A. B., PALMER P., TRAN BA HUY P., SOUDANT J., STERKERS O., LEBON P., FERRARY E. 2000. No evidence of measles virus in stapes samples from patients with otosclerosis. *J. Clin. Microbiol.* **38**: 2655 – 2660.
- GRAYELI A. B., STERKERS O., ROULLEAU P., ELBAZ P., FERRARY E., SILVE C. 1999. Parathyroid hormone-parathyroid hormone-related peptide receptor expression and function in otosclerosis. *Am. J. Physiol.* **277**: E1005 – E1012.

- GUO S. W. 1997. Linkage disequilibrium measures for fine-scale mapping: A comparison. *Hum. Hered.* **47**: 301 – 314.
- GUO S. W. & THOMPSON E. A. 1992. Performing the exact test of Hardy-Weinberg proportion for multiple alleles. *Biometrics* **48** (2): 361 – 72.
- GUSELLA J. F., WEXLER N. S., CONNEALLY P. M., NAYLOR S. L., ANDERSON M. A., TANZI R. E., WATKINS P. C., OTTINA K., WALLACE M. R., SAKAGUCHI A. Y., YOUNG A. B., SHOULSON I., BONILLA E., MARTIN J. B. 1983. A polymorphic DNA marker genetically linked to Huntington's disease. *Nature* **306** (5940): 234 – 8.
- HALDANE J. B. S. 1919. The combination of linkage values, and the calculation of distances between the loci of linked factors. *J. Genet.* **31**: 67 – 100.
- HALL D., WIJSMAN E. M., ROOS J. L., GOGOS J. A., KARAYIORGOU M. 2002. Extended intermarker linkage disequilibrium in the Afrikaners. *Genome Res.* **12** (6): 956 – 61.
- HALUSHKA M. K., FAN J. B., BENTLEY K., HSIE L., SHEN N., WEDER A., COOPER R., LIPSHUTZ R., CHAKRAVARTI A. 1999. Patterns of single-nucleotide polymorphisms in candidate genes for blood-pressure homeostasis. *Nat. Genet.* **22** (3): 239 – 47.
- HANCOCK J. M. 1996. Simple sequences in a “minimal” genome. *Nature Genet.* **14**: 14 – 15.
- HARTL D. L. & CLARK A. G. 1997. *Principles of Population Genetics*. Sinauer Associates, Sunderland.
- HASTBACKA J., DE LA CHAPELLE A., KAITILA I., SISTONEN P., WEAVER A., LANDER E. 1992. Linkage disequilibrium mapping in isolated founder populations: diastrophic dysplasia in Finland. *Nat. Genet.* **2** (3): 204 – 11.
- HASTINGS A. 1984. Linkage disequilibrium, selection and recombination at three loci. *Genetics* **106**: 153 – 164.
- HASTINGS A. 1985. Multilocus population genetics with weak epistasis. I. Equilibrium properties of two-locus two-allele models, *Genetics* **109**: 799 – 812.
- HAYES V. M. 2003. Genetic diversity of the alpha-1-antitrypsin gene in Africans identified using a novel genotyping assay. *Hum. Mut.* **22**: 59 – 66.
- HEDRICK P. W. 1987. Gametic disequilibrium measures: Proceed with caution. *Genetics* **117**: 331 – 341.
- HEDRICK P. W. 1988. Can segregation distortion influence gametic disequilibrium? *Genet. Res. Camb.* **52**: 237 – 242.
- HEDRICK P. W. 2005. *Genetics of Populations*. Jones & Bartlett Publishers, Inc.
- HEDRICK P. W., JAIN S., HOLDEN L. 1978. Multilocus systems in evolution. *Evol. Biol.* **11**: 101 – 184.
- HERMIDA J. 2001. *Desequilibrio gamético entre loci microsatélites de la región centromérica e intersticial del brazo corto del cromosoma 11 humano*. Tesis Doctoral. Universidad de Santiago de Compostela.

Bibliografía

- HERNÁNDEZ J. L. & WEIR B. S. 1989. A disequilibrium coefficient approach to Hardy-Weinberg testing. *Biometrics* **45**: 53 – 70.
- HILL W. G. 1974. Estimation of linkage disequilibrium in randomly mating populations. *Heredity* **33**: 229 – 239.
- HILL W. G. & ROBERTSON A. 1968. Linkage disequilibrium in finite populations. *Theoretical and Applied Genetics* **38**: 226 – 231.
- HOBAN P. R. & KELSEY A. M. 1991. PstI polymorphism within the 3' untranslated region of the insulin gene detectable by the polymerase chain reaction (INS). *Nuc. Acids Res.* **19**: 4576.
- HOBAN P. R., SANTIBANEZ-KOREF M. F. & KELSEY A. M. 1991. *Nuc. Acids Res.* **19**: 6976.
- HOELZEL A. R. Molecular genetic analysis of populations. Oxford University Press, 1998.
- HOUSMAN D. y LEDLEY. 1998. Why pharmacogenomics? Why now?. *Nature Biotechnology* **16**: 492 – 493.
- HOUWEN R. H. J., BAHARLOO S., BLANKENSHIP K., RAEYMAEKERS P., JUYN J., SANDKUIJL L. A., FREIMER N. B. 1994. Genome screening by searching for shared segments: mapping a gene for benign recurrent intrahepatic cholestasis. *Nature Genetics* **8**: 380 – 386.
- HUANG Q. Y., XU F. H., SHEN H., DENG H. Y., LIU Y. J., LIU Y. Z., LI J. L., RECKER R. R., DENG H. W. 2002. Mutation patterns at dinucleotide microsatellite loci in humans. *Am. J. Hum. Genet.* **70** (3): 625 – 34.
- HUTTLEY G. A., SMITH M. W., CARRINGTON M., O'BRIEN S. J. 1999. A scan for linkage disequilibrium across the human genome. *Genetics* **152** (4): 1711 – 22.
- IOANNIDIS J. P., NTZANI E. E., TRIKALINOS T. A., CONTOPOULOS-IOANNIDIS D. G. 2001. Replication validity of genetic association studies. *Nat. Genet.* **29** (3): 306 – 9.
- IOANNIDIS J. P., TRIKALINOS T. A., NTZANI E. E., CONTOPOULOS-IOANNIDIS D. G. 2003. Genetic associations in large versus small studies: an empirical assessment. *Lancet* **361** (9357): 567 – 71.
- JEFFREYS A. J., KAUPPI L. & NEUMANN R. 2001. Intensely punctuate meiotic recombination in the class II region of the major histocompatibility complex. *Nature Genetics* **29**: 217 – 222.
- JEFFREYS A. J., RITCHIE A. & NEUMANN R. 2000. High resolution analysis of haplotype diversity and meiotic crossover in the human TAP2 recombination hotspot. *Hum. Mol. Genet.* **9**: 725 – 733.
- JOHNSON G. C., ESPOSITO L., BARRATT B. J., SMITH A. N., HEWARD J., DI GENOVA G., UEDA H., CORDELL H. J., EAVES I. A., DUDBRIDGE F., TWELLS R. C., PAYNE F., HUGHES W., NUTLAND S., STEVENS H., CARR P., TUOMILEHTO-WOLF E., TUOMILEHTO J., GOUGH S. C., CLAYTON D. G., TODD J. A. 2001. Haplotype tagging for the identification of common disease genes. *Nat. Genet.* **29** (2): 233 – 7.

- JORDE L. B. 1995. Linkage disequilibrium as a gene-mapping tool. *Am. J. Hum. Genet.* **56**(1): 11 – 4.
- JORDE L. B. 2000. Linkage disequilibrium and the search for complex disease genes. *Genome Research* **10**: 1435 – 1444.
- JORDE L. B., WATKINS W. S., KERE J., NYMAN D., ERIKSSON A. W. 2000. Gene mapping in isolated populations: new roles for old friends? *Hum. Hered.* **50** (1): 57 – 65.
- JORDE L. B., WATKINS W. S., VISKOCHIL D., O'CONNELL P., WARD K. 1993. Linkage disequilibrium in the neurofibromatosis 1 (NF1) region: implications for gene mapping. *Am. J. Hum. Genet.* **53** (5): 1038 – 50.
- JORDE L. B., WATKINS W. S., CARLSON M., GRODEN J., ALBERTSEN H., THLIVERIS A., LEPPERT M. 1994. Linkage disequilibrium predicts physical distance in the adenomatous polyposis coli region. *Am. J. Hum. Genet.* **54** (5): 884 – 98.
- KAESSMANN H., HEISSIG F., VON HAESELER A., PAABO S. 1999. DNA sequence variation in a non-coding region of low recombination on the human X chromosome. *Nat. Genet.* **22**: 78 – 81.
- KAESSMANN H., ZOLLNER S., GUSTAFSSON A. C., WIEBE V., LAAN M., LUNDEBERG J., UHLEN M., PAABO S. 2002. Extensive linkage disequilibrium in small human populations in Eurasia. *Am. J. Hum. Genet.* **70** (3): 673 – 85.
- KAPLAN N. L., HUDSON R. R. & LANGLEY C. H. 1989. The "hitchhiking effect" revisited. *Genetics* **123** (4): 887 – 99.
- KARLIN S. 1969. *Equilibrium Behaviour of Population Genetics Models with Non-random Mating*. Gordon y Breach. New York.
- KASHI Y., KING D., SOLLER M. 1997. Simple sequence repeats as a source of quantitative genetic variation. *Trends Genet.* **13** (2): 74 – 8.
- KAUPPI L., SAJANTILA A., JEFFREYS A. J. 2003. Recombination hotspots rather than population history dominate linkage disequilibrium in the MHC class II region. *Hum. Mol. Genet.* **12** (1): 33 – 40.
- KE X., HUNT S., TAPPER W., LAWRENCE R., STAVRIDES G., GHORI J., WHITTAKER P., COLLINS A., MORRIS A. P., BENTLEY D., CARDON L. R., DELOUKAS P. 2004. The impact of SNP density on fine-scale patterns of linkage disequilibrium. *Hum. Mol. Genet.* **13** (6): 577 – 88.
- KENDALL F. B. A. & STUART A. 1979. *The Advanced Theory of Statistics. Vol 2. Inference and Relationship*. Charles Griffin & Company. London.
- KEREM B., ROMMENS J. M., BUCHANAN J. A., MARKIEWICZ D., COX T. K., CHAKRAVARTI A., BUCHWALD M., TSUI L. C. 1989. Identification of the cystic fibrosis gene: genetic analysis. *Science* **245** (4922): 1073 – 80.
- KEREM B. S., ZIELENSKI J., MARKIEWICZ D., BOZON D., GAZIT E., YAHAV J., KENNEDY D., RIORDAN J. R., COLLINS F. S., ROMMENS J. M. & TSUI L-C. 1990. Identification of mutations in

Bibliografia

- regions corresponding to the two putative nucleotide (ATP)-binding folds of the cystic fibrosis gene. *Proc. Natl. Acad. Sci. U S A* **87** (21): 8447 – 51.
- KIDD J. R., PAKSTIS A. J., ZHAO H., LU R. B., OKONOFUA F. E., ODUNSI A., GRIGORENKO E., TAMIR B. B., FRIEDLAENDER J., SCHULZ L. O., PARNAS J., KIDD K. K. 2000. Haplotypes and linkage disequilibrium at the phenylalanine hydroxylase locus, PAH, in a global representation of populations. *Am. J. Hum. Genet.* **66** (6): 1882 – 99.
- KIMMEL M. & CHAKRABORTY R. 1996. Measures of variation at DNA repeat loci under a general stepwise mutation model. *Theor. Popul. Biol.* **50** (3): 345 – 67.
- KIMMEL M., CHAKRABORTY R., STIVERS D. N., DEKA R. 1996. Dynamics of repeat polymorphisms under a forward-backward mutation model: within- and between-population variability at microsatellite loci. *Genetics* **143** (1): 549 – 55.
- KIMMEL M., CHAKRABORTY R., KING J. P., BAMSHAD M., WATKINS W. S., JORDE L. B. 1998. Signatures of population expansion in microsatellite repeat data. *Genetics* **148** (4): 1921 – 30.
- KING J. S., MORTIMER R. K. 1990. A polymerization model of chiasma interference and corresponding computer simulation. *Genetics* **126**: 1127 – 38.
- KOJIMA K. & SCHAFFER H. F. 1967. Survival processes of linked mutant genes. *Evolution* **21**: 518 – 531.
- KONDO M., SUZUKI H., UEDA R., OSADA H., TAKAGI K., TAKAHASHI T., TAKAHASHI T. 1995. Frequent loss of imprinting of the H19 gene is often associated with its overexpression in human lung cancers. *Oncogene* **10** (6): 1193 – 8.
- KOSAMBI D. D. 1944. The Estimation of Map Distances from Recombination Values. *Ann. Eugen. Lond.* **12**: 172 – 175.
- KRUGLYAK L., LANDER E. S. 1995. High-resolution genetic mapping of complex traits. *Am. J. Hum. Genet.* **56** (5): 1212 – 23.
- KRUGLYAK L. & NICKERSON D. A. 2001. Variation is the spice of life. *Nat. Genet.* **27**: 234 – 6.
- KRUGLYAK L. 1997. The use of a genetic map of biallelic markers in linkage studies. *Nature Genetics* **17**: 21 – 24.
- KRUGLYAK S., DURRETT R. T., SCHUG M. D., AQUADRO C. F. 1998. Equilibrium distributions of microsatellite repeat length resulting from a balance between slippage events and point mutations. *Proc. Natl. Acad. Sci. USA* **95** (18): 10774 – 8.
- KRUGLYAK L. 1999. Prospects for whole-genome linkage disequilibrium mapping of common disease genes. *Nature Genetics* **22**: 139 – 144.
- LAAN M. & PÄÄBO S. 1997. Demographic history and linkage disequilibrium in human populations. *Nature Genetics* **17** (4): 435 – 8.
- LANDER E. S. 1996. The new genomics: global views of biology. *Science* **274**: 536 – 539.
- LANDER E. S. y SCHORK N. J. 1994. Genetic dissection of complex traits. *Science* **265**: 2037 – 2048.

- LANGLEY C. H., SMITH D. B. & JHONSON F. M. 1978. Analysis of linkage disequilibria between allozyme loci in natural populations of *Drosophila melanogaster*. *Genet. Res.* **32**: 215 – 229.
- LARSSON A. 1960. Otosclerosis: a genetic and clinical study. *Acta Otolaryngol. (Stockh) (Suppl)* **154**: 1 – 86.
- LESSIOS H. A. 1992. Testing electrophoretic data for the agreement with Hardy-Weinberg expectations. *Marine Biology* **112**: 517 – 523.
- LEVINSON G., GUTMAN G. A. 1987. Slipped-strand mispairing: a major mechanism for DNA sequence evolution. *Mol. Biol. Evol.* **4** (3): 203 – 21.
- LEWONTIN R. C. 1964. The interaction of selection and linkage. I. General considerations; heterotic models. *Genetics* **49**: 49 – 67.
- LEWONTIN R. C. 1964. The interaction of selection and linkage. II. Optimum models. *Genetics* **50**: 757 – 782.
- LEWONTIN R. C. 1974. *The Genetic Basis of Evolutionary Change*. Columbia University Press. New York.
- LEWONTIN R. C. 1985. Populations Genetics. *Ann. Rev. Genet.* **19**: 81 – 102.
- LEWONTIN R. C. 1988. On Measures of Gametic Disequilibrium. *Genetics* **120**: 849 – 852.
- LEWONTIN R. C. y KOJIMA K. 1960. The evolutionary dynamics of complex polymorphisms. *Evolution* **14**: 458 – 472.
- LI W. H. & NEI M. 1974. Stable linkage disequilibrium without epistasis in subdivided populations. *Theoretical Population Biology* **6**: 173 – 183.
- LI W. H., GOJOBORI T., NEI M. 1981. Pseudogenes as a paradigm of neutral evolution. *Nature* **292**: 237 – 239.
- LI C. C. & HORVITZ D. G. 1953. Some Methods of Estimating the Inbreeding Coefficient. *Am. J. Hum. Genet.* **5** (2): 107 – 117.
- LI W. 1997. *Molecular Evolution*. Sinauer, Sunderland, MA.
- LI W. H. y SADLER L. A. 1991. Low nucleotide diversity in man. *Genetics* **129**: 513 – 523.
- LITT M. & JORDE L. B. 1986. Linkage disequilibria between pairs of loci within a highly polymorphic region of chromosome 2Q. *Am. J. Hum. Genet.* **39** (2): 166 – 78.
- LOHMUELLER K. E., PEARCE C. L., PIKE M., LANDER E. S. & HIRSCHHORN J. N. 2003. Meta-analysis of genetic association studies supports a contribution of common variants to susceptibility to common disease. *Nature Genetics* **33**: 177 – 182.
- LONG J. C., WILLIAMS R. C. & URBANEK M. 1995. An E-M algorithm and testing strategy for multiple-locus haplotypes. *Am. J. Hum. Genet.* **56**: 799 – 810.

Bibliografia

- LONGMATE J. A. 2001. Complexity and power in case-control association studies. *Am. J. Hum. Genet.* **68**: 1229 – 1237.
- LOUIS E. J. & DEMPSTER E. R. 1987. An exact test for Hardy-Weinberg and multiple alleles. *Biometrics* **43**: 805 – 811.
- LU Y. H. & NÈGRE S. 1993. Use of glycerol for enhanced efficiency and specificity of PCR amplification. *Technical Tips* **9**: 297.
- MANIATIS T., JEFFREY A., VAN DESANDE H. 1975. Chain length determination of small double- and single-stranded DNA molecules by polyacrylamide gel electrophoresis. *Biochemistry* **14** (17): 3787 – 94.
- MANIATIS N., COLLINS A., GIBSON J., ZHANG W., TAPPER W. & MORTON N. E. 2004. Positional cloning by linkage disequilibrium. *Am. J. Hum. Genet.* **74**: 846 – 855.
- MARTIN E. R., LAI E. H., GILBERT J. R., ROGALA A. R., AFSHARI A. J., RILEY J., FINCH K. L., STEVENS J. F., LIVAK K. J., SLOTTERBECK B. D., SLIFER S. H., WARREN L. L., CONNEALLY P. M., SCHMECHEL D. E., PURVIS I., PERICAK-VANCE M. A., ROSES A. D., VANCE J. M. 2000. SNPping away at complex diseases: analysis of single-nucleotide polymorphisms around APOE in Alzheimer disease. *Am. J. Hum. Genet.* **67**(2): 383 – 94.
- MARTINEZ-ARIAS R., CALAFELL F., MATEU E., COMAS D., ANDRES A., BERTRANPETIT J. 2001. Sequence variability of a human pseudogene. *Genome Res.* **11**: 1071 – 1085.
- MATISE T. C., SACHIDANANDAM R., CLARK A. G., KRUGLYAK L., WIJSMAN E., KAKOL J., BUYSKE S., CHUI B., COHEN P., DE TOMA C., EHM M., GLANOWSKI S., HE C., HEIL J., MARKIANOS K., McMULLEN I., PERICAK-VANCE M. A., SILBERGLEIT A., STEIN L., WAGNER M., WILSON A. F., WINICK J. D., WINN-DEEN E. S., YAMASHIRO C. T., CANN H. M., LAI E., HOLDEN A. L. 2003. A 3.9-centimorgan-resolution human single-nucleotide polymorphism linkage map and screening set. *Am. J. Hum. Genet.* **73** (2): 271 – 84.
- MAY C. A., SHONE A. C., KALAYDJIEVA L., SAJANTILA A. y JEFFREYS A. J. 2002. Crossover clustering and rapid decay of linkage disequilibrium in the Xp/Yp pseudoautosomal gene SHOX. *Nature Genetics* **31**: 272 – 275.
- MAYNARD-SMITH J. & HAIGH J. 1974. The *hitchhiking* Effect of a Favourable Gene. *Genet. Res.* **23**: 23 – 25.
- McKENNA M. J., KRISTIANSEN A. G., BARTLEY M. L., ROGUS J. J., HAINES J. L. Association of COL1A1 and otosclerosis: evidence for a shared genetic etiology with mild osteogenesis imperfecta. 1998. *Am. J. Otol.* **19**: 604 – 10.
- McKENNA M. J., KRISTIANSEN A. G., TROPITZSCH A. S. 2002. Similar COL1A1 expression in fibroblasts from some patients with clinical otosclerosis and those with type I osteogenesis imperfecta. *Ann. Otol. Rhinol. Laryngol.* **111**: 184 – 9.
- McPHEE J. R., GORDON M. A., RUBEN R. J., VAN DE WATER T. R. 1993. Evidence of abnormal stromelysin mRNA expression in suspected carriers of otosclerosis. A possible molecular marker. *Arch. Otolaryngol. Head Neck Surg.* **119**: 1108 – 1116.

- MENGER D. J. & TANGE R. A. 2003. The aetiology of otosclerosis: a review of the literature. *Clin. Otolaryngol.* **28**: 112 – 120.
- MILLER R. D. & KWOK P-Y. 2001. The birth and death of human single-nucleotide polymorphisms: new experimental evidence and implications for human history and medicine. *Hum. Mol. Genet.* **10**: 2195 – 2198.
- MOFFATT M. F., TRAHERNE J. A., ABECASIS G. R., COOKSON W. O. 2000. Single nucleotide polymorphism and linkage disequilibrium within the TCR alpha/delta locus. *Hum. Mol. Genet.* **9** (7): 1011 – 9.
- MOHLKE K. L., LANGE E. M., VALLE T. T., GHOSH S., MAGNUSON V. L., SILANDER K., WATANABE R. M., CHINES P. S., BERGMAN R. N., TUOMILEHTO J., COLLINS F. S., BOEHNKE M. 2001. Linkage disequilibrium between microsatellite markers extends beyond 1 cM on chromosome 20 in Finns. *Genome Res.* **11** (7): 1221 – 6.
- MORGANTE M., HANAFEY M. & POWELL W. 2002. Microsatellites are preferentially associated with nonrepetitive DNA in plant genomes. *Nature Genet.* **30**: 194 – 200.
- MORRISON A. W. 1967. Genetic factors in otosclerosis. *Annals of the Royal College of Surgeons of England.* **41**: 202 – 237.
- MORTIMER R. K. & FOGEL S. 1974. Genetical interference and gene conversion. *Mechanisms in Recombination*. Plenum, New York.
- MORTON N. E., COLLINS A., LAWRENCE S., SHIELDS D. C. 1992. Algorithms for a location database. *Ann. Hum. Genet.* **56**: 223 – 32.
- MULLER H. J. 1916. The mechanism of crossing over. II. *Amer. Naturalist* **50**: 284 – 305.
- MULLERMAN J. E., SHIELDS J. J. & SAHA B. K. 1992. Characterization of two novel polymorphisms at the human parathyroid hormone gene locus. *Hum. Genet.* **88**: 589 – 592.
- MULLIS K. B., FALOONA F., SCHARF S., SAIKI R., HORN G. & ERLICH H. 1986. Specific Enzymatic Amplification of DNA “in vitro”, the Polymerase Chain Reaction. *Cold Spring Harb. Symp. Quant. Biol.* **51** Pt 1: 263 – 273.
- NACHMAN M. W. 2001. Single nucleotide polymorphisms and recombination rate in humans. *Trends in Genetics.* **17**: 481 – 485.
- NACHMAN M. W., BAUER V. L., CROWELL S. L. y AQUADRO C. F. 1998. DNA variability and recombination rates at X-linked loci in humans. *Genetics* **150**: 1133 – 1141.
- NAGAI M., YOSHIDA A. & SATO N. 1998. Additive effects of bovine serum albumin, dithiothreitol, and glycerol on PCR. *Biochem. Mol. Biol. Int.* **44** (1): 157 – 63.
- NAKAJIMA T., JORDE L. B., ISHIGAMI T., UMEMURA S., EMI M., LALOUEL J. M., INOUE I. 2002. Nucleotide diversity and haplotype structure of the human angiotensinogen gene in two populations. *Am. J. Hum. Genet.* **70** (1): 108 – 23.
- NAUTA M. J., WEISSING F. J. 1996. Constraints on allele size at microsatellite loci: implications for genetic differentiation. *Genetics* **143** (2): 1021 – 32.

Bibliografía

- NEALE B. M. & SHAM P. C. 2004. The future of association studies: Gene-based analysis and replication. *Am. J. Hum. Genet.* **75**: 353 – 362.
- NEI M. 1978. Estimation of Average Heterozygosity and Genetic Distance from a Small Number of Individuals. *Genetics* **89**: 583 – 590.
- NEI M. & LI W. H. 1973. Linkage disequilibrium in subdivided populations. *Genetics* **75**: 213 – 219.
- NEI M. & ROYCHOUDHURY A. K. 1974. Sampling Variances of Heterozygosity and Genetic Distance. *Genetics* **76**: 379 – 390.
- NIEDERMEYER H. P., ARNOLD W. 1995. Otosclerosis: A measles virus associated immune response of the inner ear. *Acta Otolaryngol. (Stockh)*. **115**: 300 – 303.
- NORDBORG M. y TAVARÉ S. 2002. Linkage disequilibrium: what history has to tell us. *Trends Genetics* **18**: 83 – 90.
- NOWAK R. 1994. Mining treasures from 'junk DNA'. *Science* **263** (5147): 608 – 10.
- OGILVIE R. F. & HALL I. S. 1962. On the aetiology of otosclerosis. *J. Laryngol. Otol.* **76**: 841 – 896.
- OHASHI J. & TOKUNAGA K. 2003. Power of genome-wide linkage disequilibrium testing by using microsatellite markers. *J. Hum. Genet.* **48**: 487 – 491.
- OHATA T. & KIMURA M. 1969. Linkage disequilibrium due to the random genetic drift. *Genet. Res.* **13**: 47 – 55.
- OHATA T. & KIMURA M. 1973. The Model of Mutation Appropriate to Estimate the Number of Electrophoretically Detectable Alleles in Genetic Population. *Genet. Res.* **22**: 201 – 204.
- OHATA T. & KIMURA M. 1975. The effect of selected linked locus on heterozygosity of neutral alleles (the hitch-hiking effect). *Genet. Res.* **25**: 313 – 326.
- OHATA T. 1982. Linkage disequilibrium with the island model. *Genetics* **101**: 139-155.
- OHATA T. 1982. Linkage disequilibrium due to random genetic drift in finite subdivided populations. *Proc. Natl. Acad. Sci. USA* **79**: 1940 – 1944.
- OTT J. & RABINOWITZ D. 1997. The effect of marker heterozygosity on the power to detect linkage disequilibrium. *Genetics* **147**: 927 – 930.
- OTT J. 1999. *Analysis of Human Genetic Linkage*. The Johns Hopkins University Press, Baltimore.
- OTT J. 2000. Predicting the range of linkage disequilibrium. *Proc. Natl. Acad. Sci. USA* **97**: 2 – 3.
- PÄÄBO S. 2003. The mosaic that is our genome. *Nature* **421**: 409 – 412.

- PAOLONI-GIACOBINO A., MOUTHON D., LAMBERCY C., VESSAZ M., COUTANT-ZIMMERLI S., RUDOLPH W., MALAFOSSE A., BURESI C. 2000. Identification and analysis of new sequence variants in the human tryptophan hydroxylase (TpH) gene. *Mol. Psychiatry* **5** (1): 49 – 55.
- PATIL N., BERNO A. J., HINDS D. A., BARRETT W. A., DOSHI J. M., HACKER C. R., KAUTZER C. R., LEE D. H., MARJORIBANKS C., McDONOUGH D. P., NGUYEN B. T., NORRIS M. C., SHEEHAN J. B., SHEN N., STERN D., STOKOWSKI R. P., THOMAS D. J., TRULSON M. O., VYAS K. R., FRAZER K. A., FODOR S. P., COX D. R. 2001. Blocks of limited haplotype diversity revealed by high-resolution scanning of human chromosome 21. *Science* **294** (5547): 1719 – 23.
- PATTERSON N., HATTANGADI N., LANE B., LOHMEUILLER K. E., HAFNER D. A., OKSENBURG J. R., HAUSER S. L., SMITH M. W., O'BRIEN S. J., ALTSHULER D., DALY M. J., REICH D. 2004. Methods for high-density admixture mapping of disease genes. *Am. J. Hum. Genet.* **74** (5): 979 – 1000.
- PAYSEUR B. A., NACHMAN M. W. 2000. Microsatellite variation and recombination rate in the human genome. *Genetics* **156** (3): 1285 – 98.
- PEDERSON U., MELSEN F., ELBROND O., CHARLES P. 1985. Histopathology of the stapes in osteogenesis. *J. Laryngol. Otol.* **99**: 451 – 458.
- PEPE G. 1993. A highly polymorphic (ACT)_n VNTR (variable number of tandem repeats) locus inside intron 12 of COL1A2, one of the two genes involved in fominant osteogenesis imperfecta. *Hum. Mut.* **2**: 300 – 305.
- PEPE G., RICKARDS O., JODICE C., MODIANO G. 1995. Allele and haplotype frequency distribution of the EcoRI, RsaI, and MspI COL1A2 RFLPs among various human populations. *Hum. Biol.* **67** (6): 905 – 20.
- PETERSEN D. C., LATEN A., ZEIER M. D., GRIMWOOD A., RENSBERG E. J., HAYES V. M. 2002. Novel mutations and SNPs identified in CCR2 using a new comprehensive denaturing gradient gel electrophoresis assay. *Hum. Mutat.* **20** (4): 253 – 9.
- PETERSON A. C., DI RIENZO A., LEHESJOKI A. E., DE LA CHAPPELLE A., SLATKIN M., FREIMER N. B. 1995. The distribution of linkage disequilibrium over anonymous genome regions. *Hum. Mol. Genet.* **4** (5): 887 – 94.
- PETRONIS A., VAN TOL H. H. M. & KENNEDY J. L. 1994. A SmaI PCR-RFLP in the 5' noncoding region of the human D4 Dopamin Receptor gene (DRD4). *Hum. Hered.* **44**: 58 – 60.
- PHILLIPS M. S., LAWRENCE R., SACHIDANANDAM R., MORRIS A. P., BALDING D. J., DONALDSON M. A., STUDEBAKER J. F., ANKENER W. M., ALFISI S. V., KUO F. S., CAMISA A. L., PAZOROV V., SCOTT K. E., CAREY B. J., FAITH J., KATARI G., BHATTI H. A., CYR J. M., DEROHANNESSIAN V., ELOSUA C., FORMAN A. M., GRECCO N. M., HOCK C. R., KUEBLER J. M., LATHROP J. A., MOCKLER M. A., NACHTMAN E. P., RESTINE S. L., VARDE S. A., HOZZA M. J., GELFAND C. A., BROXHOLME J., ABECASIS G. R., BOYCE-JACINO M. T., CARDON L. R. 2003. Chromosome-wide distribution of haplotype blocks and the role of recombination hot spots. *Nature Genetics* **33**: 382 – 387.

Bibliografía

- POLLOCK D. D., BERGMAN A., FELDMAN M. W., GOLDSTEIN D. B. 1998. Microsatellite behavior with range constraints: parameter estimation and improved distances for use in phylogenetic reconstruction. *Theor. Popul. Biol.* **53** (3): 256 – 71.
- PRIMMER C. R., ELLEGREN H. 1998. Patterns of molecular evolution in avian microsatellites. *Mol. Biol. Evol.* **15** (8): 997 – 1008.
- PRIMMER C. R., SAINO N., MOLLER A. P., ELLEGREN H. 1996. Directional evolution in germline microsatellite mutations. *Nat. Genet.* **13** (4): 391 – 3.
- PRINZ M., SCHMITT C., STAAK M., BAUM H. 1996. Resolution of apolipoprotein B repeat unit position variants on agarose, denaturing, and native polyacrylamide gels. *Electrophoresis* **17** (7): 1190 – 3.
- PRITCHARD J. K. 2001. Are rare variants responsible for susceptibility to complex diseases?. *Am. J. Hum. Genet.* **69** (1): 124 – 37.
- PRITCHARD J. K. & COX N. J. 2002. The allelic architecture of human disease genes: common disease-common variant...or not?. *Hum. Mol. Genet.* **11**(20): 2417 – 23.
- PRITCHARD J. K. y PRZEWORSKI M. 2001. Linkage disequilibrium in humans: Models and data. *Am. J. Hum. Genet.* **69**: 1 – 14.
- PRITCHARD J. K. & ROSENBERG N. A. 1999. Use of unlinked genetic markers to detect population stratification in association studies. *Am. J. Hum. Genet.* **65** (1): 220 – 8.
- PROCKOP D. J. 1990. Mutations that alter the primary structure of type I collagen. The perils of a system for generating large structures by the principle of nucleated growth. *J. Biol. Chem.* **265**: 15349 – 52.
- PROCKOP D. J., KIVIRIKKO K. I. 1995. Collagens: molecular biology, diseases, and potentials for therapy. *Annu. Rev. Biochem.* **64**: 403 – 34.
- PRZEWORSKI M. 2002. The signature of positive selection at randomly chosen loci. *Genetics* **160**: 1179 – 1189.
- PRZEWORSKI M., HUDSON R. R., DI RIENZO A. 2000. Adjusting the focus on human variation. *Trends Genet.* **16**: 296 – 302.
- PTAK S. E., VOELPEL K. & PRZEWORSKI M. 2004. Insights into recombination from patterns of linkage disequilibrium in humans. *Genetics* **167**: 387 – 397.
- RAMEL C. 1997. Mini- and microsatellites. *Env. Health Persp.* **105**: 781 – 789.
- RAYMOND M. & ROUSSET F. 1995. An exact test for population differentiation. *Evolution* **49**: 1280 – 1283.
- REDFORD-BADWAL D. A., STOVER M. L., VALLI M., MCKINSTRY M. B., ROWE D. W. 1996. Nuclear retention of COL1A1 messenger RNA identifies null alleles causing mild osteogenesis imperfecta. *J. Clin. Invest.* **97**: 1035 – 40.

- REICH D. E., CARGILL M., BOLK S., IRELAND J., SABETI P. C., RICHTER D. J., LAVERY T., KOUYOUJIAN R., FARHADIAN S. F., WARD R., LANDER E. S. 2001. Linkage disequilibrium in the human genome. *Nature* **411**: 199 – 204.
- REICH D. E., SCHAFFNER S. F., DALY M. J., McVEAN G., MULLIKIN J. C., HIGGINS J. M., RICHTER D. J., LANDER E. S. & ALTSHULER D. 2002. Human genome sequence variation and the influence of gene history, mutation and recombination. *Nature Genetics* **32**: 135 – 142.
- RICE W. R. 1989. Analysing tables of statistical tests. *Evolution* **43** (1): 223 – 225.
- RICHIE K. L., GOLDSBOROUGH M. D., DARFLER M. M., BENZINGER E. A., LOVEKAMP M. L., REEDER D. J., O'CONNELL C. D. 1999. Long PCR for VNTR analysis. *J. Forensic. Sci.* **44** (6): 1176 – 85.
- RIEDER M. J., TAYLOR S. L., CLARK A. G., NICKERSON D. A. 1999. Sequence variation in the human angiotensin converting enzyme. *Nat. Genet.* **22** (1): 59 – 62.
- RIOUX J. D., DALY M. J., SILVERBERG M. S., LINDBLAD K., STEINHART H., COHEN Z., DELMONTE T., KOCHER K., MILLER K., GUSCHWAN S., KULBOKAS E. J., O'LEARY S., WINCHESTER E., DEWAR K., GREEN T., STONE V., CHOW C., COHEN A., LANGELIER D., LAPOINTE G., GAUDET D., FAITH J., BRANCO N., BULL S. B., MCLEOD R. S., GRIFFITHS A. M., BITTON A., GREENBERG G. R., LANDER E. S., SIMINOVITCH K. A., HUDSON T. J. 2001. Genetic variation in the 5q31 cytokine gene cluster confers susceptibility to Crohn disease. *Nat. Genet.* **29** (2): 223 – 8.
- RISCH N. 2000. Searching for genetic determinants in the new millennium. *Nature* **405** (6788): 847 – 56.
- RISCH N. y MERIKANGAS K. 1996. The future of genetic studies of complex human diseases. *Science* **273**: 1516 – 1517.
- ROBERTSON A. & HILL W. G. 1984. Deviations from Hardy-Weinberg Proportions: Sampling Variances and Use in Estimation of Inbreeding Coefficients. *Genetics* **107**: 703 – 718.
- RODRÍGUEZ L., HERMIDA J., SANDE E., RODRÍGUEZ S., NÚÑEZ C., LOURO C., MARTÍN C. & ZAPATA C. 2003. Marker-dependent background disequilibrium on human chromosome 11p. *Am. J. Hum. Genet.* **73** (Supl): 488.
- RODRÍGUEZ L., RODRÍGUEZ S., HERMIDA J., FRADE C., SANDE E., VISEDO G., MARTÍN C. & ZAPATA C. 2004. Proposed association between the COL1A1 and COL1A2 genes and otosclerosis is not supported by a case-control study in Spain. *Am. J. Hum. Genet.* **128A**: 19 – 22.
- RODRÍGUEZ S., VISEDO G. & ZAPATA C. 2001. Detection of Errors in Dinucleotide Repeat Typing by Nondenaturing Electrophoresis. *Electrophoresis* **22**: 2656 – 2664.
- RODRÍGUEZ S. & ZAPATA C. 2002. Typing Dinucleotide Repeats Under Nondenaturing Conditions UIT Single-Base Resolution and High Sizing Precision. *Molecular Biotechnology* **21**: 117 – 122.

Bibliografía

- ROHLF F. J. 1993. NTSYS-pc. Numerical Taxonomy and Multivariate Analysis System. Exeter, Setauket, New York.
- ROSES A. D. 2001. Pharmacogenetics. *Human Molecular Genetics* **10**: 2261 – 2267.
- ROTHMAN K. J. 1990. No Adjustments are Need for Multiple Comparisons. *Epidemiology* **1**: 43 – 46.
- ROUSSET F. & RAYMOND M. 1995. Testing heterozygote excess and deficiency. *Genetics*. **140** (4): 1413 – 9.
- RYBICKI B. A., IYENGAR S. K., HARRIS T., LIPTAK R., ELSTON R. C., SHEFFER R., CHEN K. M., MAJOR M., MALIARIK M. J., IANNUZZI M. C. 2002. The distribution of long range admixture linkage disequilibrium in an African-American population. *Hum. Hered.* **53** (4): 187 – 96.
- SABETI P. C., REICH D. E., HIGGINS J. M., LEVINE H. Z., RICHTER D. J., SCHAFFNER S. F., GABRIEL S. B., PLATKO J. V., PATTERSON N. J., McDONALD G. J., ACKERMAN H. C., CAMPBELL S. J., ALTSHULER D., COOPER R., KWIAKOWSKI D., WARD R., LANDER E. S. 2002. Detecting recent positive selection in the human genome from haplotype structure. *Nature* **419**: 832 – 837.
- SAIKI R. K., SCHARF S., FALOONA F., MULLIS K. B., HORN G. T., ERLICH H. A., ARNHEIM N. 1985. Enzymatic amplification of β -globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. *Science* **230**: 1350 – 1354.
- SALAS A., COMAS D., LAREU M. V., BERTRANPETIT J. & CARRACEDO A. 1998. mtDNA Analysis of the Galician Population: a Genetic Edge of European Variation. *Eur. J. Hum. Genet.*, **6**: 365 – 375.
- SANCHEZ G., GAUTHERET D., XU X., CHENINE A. L., HIRSCH I. 1998. Relative amplification efficiency of differently sized templates by long-distance PCR. *Biotechniques* **24** (3): 400 – 2.
- SANDE E. 2002. *Desequilibrio gamético entre loci microsatélites de la región telomérica del brazo largo del cromosoma X humano*. Tesis Doctoral. Universidad de Santiago de Compostela.
- SANDE E., HERMIDA J., RODRÍGUEZ L., RODRÍGUEZ S. & ZAPATA C. 2003. Patterns of gametic disequilibrium between microsatellites loci on human Xq and 11p chromosomes. *Am. J. Hum. Genet.* **73** (Supl): 388.
- SCHAI D. J. & SOMMER S. S. 1993. Genotype relative risks: methods for design and analysis of candidate-gene association studies. *Am. J. Hum. Genet.* **53**(5): 1114 – 26.
- SCHLÖTTERER C. en: Hoelzel, A. R. *Molecular Genetic Analysis of Populations: A Practical Approach*, Oxford University Press, New York 1998, pp. 237 – 261.
- SCHLÖTTERER C. 2000. Evolutionary dynamics of microsatellite DNA. *Chromosoma* **109**: 365 – 371.

- SCHLÖTTERER C., RITTER R. HARR B. & BREM G. 1998. High mutation rate of a long microsatellite allele in *Drosophila melanogaster* provides evidence for allele-specific mutation rates. *Mol. Biol. Evol.* **15**: 1269 – 1274.
- SCHLOTTERER C., TAUTZ D. 1992. Slippage synthesis of simple sequence DNA. *Nucleic Acids Res.* **20** (2): 211 – 5.
- SCHNEIDER S., KUEFFER J. M., ROESSLI K., EXCOFFIER L. 2000. Arlequin ver. 2000: a software for population genetic data analysis. Genetics and Biometry Lab., Univ. Geneva, Switzerland.
- SCHORK N. J. 1997. Genetics of complex disease. Approaches, problems and solutions. *Am. J. Respir. Crit. Care. Med.* **156**: S103 – S109.
- SCHULZE T. G., CHEN Y. S., AKULA N., HENNESSY K., BADNER J. A., MCINNIS M. G., DEPAULO J. R., SCHUMACHER J., CICHON S., PROPPING P., MAIER W., RIETSCHER M., NOTHEN M. M., McMAHON F. J. 2002. Can long-range microsatellite data be used to predict short-range linkage disequilibrium? *Hum. Mol. Genet.* **11** (12): 1363 – 72.
- SCHULZE T. G., ZHANG K., CHEN Y. S., AKULA N., SUN F., McMAHON F. J. 2004. Defining haplotype blocks and tag single-nucleotide polymorphisms in the human genome. *Hum. Mol. Genet.* **13** (3): 335 – 42.
- SEMENZA G. L., DOWLING C. E. & KAZAZIAN H. H. 1989. Hinf I polymorphism 3' to the human β -globin gene detected by the polymerase chain reaction (PCR). *Nuc. Acids Res.* **17**: 1989.
- SERVICE S. K., OPHOFF R. A., FREIMER N. B. 2001. The genome-wide distribution of background linkage disequilibrium in a population isolate. *Hum. Mol. Genet.* **10** (5): 545 – 51.
- SHAM P. C., ZHAO J. H. & CURTIS D. 2000. The effect of marker characteristics on the power to detect linkage disequilibrium due to single or multiple ancestral mutations. *Ann. Hum. Genet.* **64**: 161 – 169.
- SHEFFIELD V. C., COX D. R., LERMAN L. S., MYERS R. M. 1989. Attachment of a 40-base-pair G + C-rich sequence (GC-clamp) to genomic DNA fragments by the polymerase chain reaction results in improved detection of single-base changes. *Proc. Natl. Acad. Sci. U S A* **86** (1): 232 – 6.
- SHEFFIELD V. C., STONE E. M., CARMI R. 1998. Use of isolated inbred human populations for identification of disease genes. *Trends Genet.* **14**: 391 – 396.
- SHIFMAN S., KUYPERS J., KOKORIS M., YAKIR B., DARVASI A. 2003. Linkage disequilibrium patterns of the human genome across populations. *Hum. Mol. Genet.* **12** (7): 771 – 6.
- SHRIVER M. D., JIN L., CHAKRABORTY R., BOERWINKLE E. 1993. VNTR allele frequency distributions under the stepwise mutation model: a computer simulation approach. *Genetics* **134** (3): 983 – 93.
- SHRIVER M. D., JIN L., FERRELL R. E., DEKA R. 1997. Microsatellite data support an early population expansion in Africa. *Genome Res.* **7** (6): 586 – 91.

Bibliografia

- SINGHAL S. K., DLO, MS, DIPNBE, SHER BAJ SINGH MANN, MS, MNAMS, USHA DATTA, MD, MNAMS, FIC(PATH), NARESH K. PANDA, MS, DIPNBE, ASHOK KUMAR GUPTA, DLO, MS. 1999. Genetic correlation in otosclerosis. *Am. J. Otolaryngol.* **20**: 102 – 105.
- SINNOCK P. & SING C. F. 1972. Analysis of multilocus genetic systems in Tecumseh, Michigan. II. Consideration of the correlation between nonalleles in gametes. *Am. J. Hum. Genet.* **24**: 393 – 415.
- SLATKIN M. 1972. On treating the chromosome as the unit of selection. *Genetics* **72**: 157 – 168.
- SLATKIN M. 1994. Linkage disequilibrium in growing and stable populations. *Genetics* **137**: 331 – 336.
- SMOUSE P. E. & NEEL J. V. 1977. Multivariate analysis of gametic disequilibrium in the yanomama. *Genetics* **85**: 733 – 752.
- SOKOLOV B. P., PRYTKOV A. N., TROMP G., KNOWLTON R. G., PROCKOP D. J. 1991. Exclusion of COL1A1, COL1A2, and COL3A1 genes as candidate genes for Ehlers-Danlos syndrome type I in one large family. *Hum. Genet.* **88** (2): 125 – 9.
- SPIELMAN R. S. & EWENS W. J. 1996. The TDT and other family-based tests for linkage disequilibrium and association. *Am. J. Hum. Genet.* **59**(5): 983 – 9.
- STAHL F. W. 1979. Genetic Recombination: Thinking About it in Phage and Fungi. Freeman and Company. San Francisco. USA.
- STEPHAN W. & KIM Y. 1998. Persistence of microsatellite arrays in finite populations. *Mol. Biol. Evol.* **15** (10): 1332 – 6.
- STEPHENS J. C., BRISCOE D. & O'BRIEN S. J. 1994. Mapping by admixture linkage disequilibrium in human populations: Limits and guidelines. *Am. J. Hum. Genet.* **55**: 809 – 824.
- STEPHENS J. C., SCHNEIDER J. A., TANGUAY D. A., CHOI J., ACHARYA T., STANLEY S. E., JIANG R., MESSER C. J., CHEW A., HAN J. H., DUAN J., CARR J. L., LEE M. S., KOSHY B., KUMAR A. M., ZHANG G., NEWELL W. R., WINDEMUTH A., XU C., KALBFLEISCH T. S., SHANER S. L., ARNOLD K., SCHULZ V., DRYSDALE C. M., NANDABALAN K., JUDSON R. S., RUANO G., VOVIS G. F. 2001. Haplotype variation and linkage disequilibrium in 313 human genes. *Science* **293** (5529): 489 – 93.
- STRACHAN T. & READ A. P. Human Molecular Genetics 2. BIOS Scientific publishers, 1999.
- STUMPF M. P. H. & GOLDSTEIN D. B. 2003. Demography, recombination hotspot intensity and the block structure of linkage disequilibrium. *Curr. Biol.* **13**: 1 – 8.
- STURTEVANT A. H. 1915. The behaviour of chromosomes as studied through linkage. *Indukt. Z. Abstammungs. Vererbungsl.* **13**: 234 – 287.
- SVED J. A. 1971. Linkage disequilibrium and homozygosity of chromosome segments in finite populations. *Theor. Popul. Biol.* **2** (2): 125 – 41.

- SYKES B., OGILVIE D., WORDSWORTH P., WALLIS G., MATHEW C., BEIGHTON P., NICHOLLS A., POPE F. M., THOMPSON E., TSIPOURAS P., SCHWARTZ R., JENSSON O., ARNASON A., BORRESEN A-L., HEIBERG A., FREY D., STEINMANN B. 1990. Consistent linkage of dominantly inherited osteogenesis imperfecta to the type I collagen loci: COL1A1 and COL1A2. *Am. J. Hum. Genet.* **46**: 293 – 307.
- TADOKORO K., OKI N., SAKAI A., FUJII H., OHSHIMA A., NAGAFUCHI S., INOUE T., YAMADA M. 1993. PCR detection of 9 polymorphisms in the WT1 gene. *Hum. Mol. Genet.* **2** (12): 2205 – 6.
- TAILLON-MILLER P., BAUER-SARDINA I., SACCONI N. L., PUTZEL J., LAITINEN T., CAO A., KERE J., PILIA G., RICE J. P., KWOK P. Y. 2000. Juxtaposed regions of extensive and minimal linkage disequilibrium in human Xq25 and Xq28. *Nat. Genet.* **25** (3): 324 – 8.
- TANAKA G., MATSUSHITA I., OHASHI J., TSUCHIYA N., IKUSHIMA S., ORITSU M., HIJIKATA M., NAGATA T., YAMAMOTO K., TOKUNAGA K., KEICHO N. 2005. Evaluation of microsatellite markers in association studies: a search for an immune-related susceptibility gene in sarcoidosis. *Immunogenetics* **56** (12): 861 – 70.
- TAPAROWSKY E., SUARD Y., FASANO O., SHIMIZU K., GOLDFARB M., WIGLER M. 1982. Activation of the T24 bladder carcinoma transforming gene is linked to a single amino acid change. *Nature* **300** (5894): 762 – 5.
- TEMPLETON A. R., CLARK A. G., WEISS K. M., NICKERSON D. A., BOERWINKLE E. & SING C. F. 2000. Recombinational and mutational hotspots within the human lipoprotein lipase gene. *Am. J. Hum. Genet.* **66**: 69 – 83.
- TERWILLIGER J. D. 1995. A powerful likelihood method for the analysis of linkage disequilibrium between trait loci and one or more polymorphic marker loci. *Am. J. Hum. Genet.* **56**(3): 777 – 87.
- TERWILLIGER J. D. & GÖRING H. H. 2000. Gene mapping in the 20th and 21st centuries: statistical methods, data analysis, and experimental design. *Hum Biol.* **72**(1): 63 – 132.
- TERWILLIGER J.D. & WEISS K.M. 1998. Linkage disequilibrium mapping of complex disease: fantasy or reality?. *Curr. Opin. Biotechnol.* **9**(6): 578 – 94.
- TERWILLIGER J. D., ZOLLNER S., LAAN M., PAABO S. 1998. Mapping genes through the use of linkage disequilibrium generated by genetic drift: 'drift mapping' in small populations with no demographic expansion. *Hum. Hered.* **48** (3): 138 – 54.
- THE INTERNATIONAL HAPMAP CONSORTIUM. 2003. The International HapMap Project. *Nature* **426**: 789 – 796.
- THOMPSON E. A., DEEB S., WALKER D. & MOTULSKY A. G. 1988. The detection of linkage disequilibrium between closely linked markers: RFLPS at the AI-CIII Apolipoprotein genes. *Am. J. Hum. Genet.* **42**: 113 – 124.
- THOMSON G. 2001. Significance levels in genome scans. *Adv. Genet.* **42**: 475 – 486.
- TIRET L., POIRIER O., NICAUD V., BARBAUX S., HERRMANN S. M., PERRET C., RAOUX S., FRANCOMME C., LEBARD G., TREGOUET D. & CAMBIEN F. Heterogeneity of linkage

Bibliografia

- disequilibrium in human genes has implications for association studies of common diseases. 2002. *Hum. Mol. Genet.* **11**: 419 – 429.
- TISHKOFF S. A., DIETZSCH E., SPEED W., PAKSTIS A. J., KIDD J. R., CHEUNG K., BONNÉ-TAMIR B., BENERECETTI A. S. S., MORAL P. KRINGS M., PÄÄBO S., WATSON E., RISCH N., JENKINS T., KIDD K. K. 1996. Global patterns of linkage disequilibrium at the CD4 locus and modern human origins. *Science* **271**: 1380 – 1387.
- TISHKOFF S. A., GOLDMAN A., CALAFELL F., SPEED W. C., DEINARD A. S., BONNE-TAMIR B., KIDD J. R., PAKSTIS A. J., JENKINS T., KIDD K. K. 1998. A global haplotype analysis of the myotonic dystrophy locus: implications for the evolution of modern humans and for the origin of myotonic dystrophy mutations. *Am. J. Hum. Genet.* **62** (6): 1389 – 402.
- TISHKOFF S. A., PAKSTIS A. J., STONEKING M., KIDD J. R., DESTRO-BISOL G., SANJANTILA A., LU R. B., DEINARD A. S., SIRUGO G., JENKINS T., KIDD K. K., CLARK A. G. 2000. Short tandem-repeat polymorphism/alu haplotype variation at the PLAT locus: implications for modern human origins. *Am. J. Hum Genet.* **67**(4): 901 – 25.
- TISHKOFF S. A., VARKONYI R., CAHINHINAN N., ABBES S., ARGYROPOULOS G., DESTRO-BISOL G., DROUSIOTOU A., DANGERFIELD B., LEFRANC G., LOISELET J., PIRO A., STONEKING M., TAGARELLI A., TAGARELLI G., TOUMA E. H., WILLIAMS S. M., CLARK A. G. 2001. Haplotype diversity and linkage disequilibrium at human GSPD: recent origin of alleles that confer malarial resistance. *Science* **293**: 455 – 462.
- TODD J. A. 1999. Interpretation of results from genetic studies of multifactorial diseases. *Lancet* **354**: 15 – 16.
- TOMEK M. S., BROWN M. R., MANI S. R., RAMESH A., SRISAILAPATHY C. R. S., COUCKE P., ZBAR R. I. S., BELL A. M., MCGUIRT W. T., FUKUSHIMA K., WILLEMS P. J., VAN CAMP G., SMITH R. J. H. 1998. Localization of a gene for otosclerosis to chromosome 15q25-q26. *Hum. Mol. Genet.* **7**: 285 – 290.
- TÓTH G., GASPARI Z. & JURKA J. 2000. Microsatellites in different eukaryotic genomes: survey and analysis. *Genome Res.* **10**: 967 – 981.
- TSUNODA T., LATHROP G. M., SEKINE A., YAMADA R., TAKAHASHI A., OHNISHI Y., TANAKA T., NAKAMURA Y. 2004. Variation of gene-based SNPs and linkage disequilibrium patterns in the human genome. *Hum. Mol. Genet.* **13** (15): 1623 – 32.
- TWELLS R. C., MEIN C. A., PHILLIPS M. S., HESS J. F., VEIJOLA R., GILBEY M., BRIGHT M., METZKER M., LIE B. A., KINGSNORTH A., GREGORY E., NAKAGAWA Y., SNOOK H., WANG W. Y., MASTERS J., JOHNSON G., EAVES I., HOWSON J. M., CLAYTON D., CORDELL H. J., NUTLAND S., RANCE H., CARR P. & TODD J. A. 2003. Haplotype structure, LD blocks, and uneven recombination within the LRP5 gene. *Genome Res.* **13**: 845 – 855.
- UPTON G. J. G. 1982. A comparison of alternative tests for the 2 × 2 comparative trial. *J. R. Statist. Soc. A.* **145**: 86 – 105.
- VALDES A. M., SLATKIN M., FREIMER N. B. 1993. Allele frequencies at microsatellite loci: the stepwise mutation model revisited. *Genetics* **133** (3): 737 – 49.
- VAN DEN BOGAERT K., GOVAERTS P. J., DE LEENHEER E. M. R., SCHATTEMAN I., VERSTREKEN M., CHEN W., DECLAU F., CREMERS C. W. R. J., VAN DE HEYNING P. H., OFFECIERS F. E.,

- SOMERS T., SMITH R. J. H., VAN CAMP G. 2002. Otosclerosis: A genetically heterogeneous disease involving at least three different genes. *Bone* **30**: 624 – 630.
- VAN DEN BOGAERT K., GOVAERTS P. J., SCHATTEMAN I., BROWN M. R., CAETHOVEN G., OFFECIERS F., SOMERS T., SECLAU F., COUCKE P., VAN DE HEYNING P., SMITH R. J. H., VAN CAMP G. 2001. A second gene for otosclerosis, OTSC2, maps to chromosome 7q34-36. *Am. J. Hum. Genet.* **68**: 495 – 500.
- VAN DEN BOGAERT K., DE LEENHEER E. M., CHEN W., LEE Y., NURNBERG P., PENNING S. R. J., VANDERSTRAETEN K., THYS M., CREMERS C. W., SMITH R. J., VAN CAMP G. 2004. A fifth locus for otosclerosis, OTSC5, maps to chromosome 3q22-24. *J. Med. Genet.* **41** (6): 450 – 3.
- VARILO T., PAUNIO T., PARKER A., PEROLA M., MEYER J., TERWILLIGER J. D. & PELTONEN L. 2003. The interval of linkage disequilibrium (LD) detected with microsatellite and SNP markers in chromosomes of Finnish populations with different histories. *Hum. Mol. Genet.* **12**: 51 – 59.
- VARILO T., PELTONEN L. 2004. Isolates and their potential use in complex gene mapping efforts. *Curr. Opin. Genet. Dev.* **14** (3): 316 – 23.
- VIELAND V. J. 2001. The replication requirement. *Nat. Genet.* **29**: 244 – 245.
- VIGNAL A., MILAN D., SANCRISTOBAL M., EGGEN A. 2002. A review on SNP and other types of molecular markers and their use in animal genetics. *Genet. Sel. Evol.* **34**: 275 – 305.
- WALL J. D. 2004. Close look at gene conversion hot spots. *Nat. Genet.* **36**: 114 – 115.
- WALL J. D. y PRITCHARD J. K. 2003a. Haplotype blocks and linkage disequilibrium in the human genome. *Nature Reviews Genetics* **4**: 587 – 597.
- WALL J. D. y PRITCHARD J. K. 2003b. Assessing the performance of the haplotype block model of linkage disequilibrium. *Am. J. Hum. Genet.* **73**: 502 – 515.
- WALL J. D., FRISSE L. A., HUDSON R. R. & DI RIENZO A. 2003. Comparative linkage-disequilibrium analysis of the β -globin hotspot in primates. *Am. J. Hum. Genet.* **73**: 1330 – 1340.
- WANG D. G., FAN J. B., SIAO C. J., BERNO A., YOUNG P., SAPOLSKY R., GHANDOUR G., PERKINS N., WINCHESTER E., SPENCER J., KRUGLYAK L., STEIN L., HSIE L., TOPALOGLOU T., HUBBELL E., ROBINSON E., MITTMANN M., MORRIS M. S., SHEN N., KILBURN D., RIOUX J., NUSBAUM C., ROZEN S., HUDSON T. J., LIPSHUTZ R., CHEE M., LANDER E. S. 1998. Large-scale identification, mapping, and genotyping of single-nucleotide polymorphisms in the human genome. *Science* **280** (5366): 1077 – 82.
- WANG N., AKEY J. M., ZHANG K., CHAKRABORTY R., JIN L. 2002. Distribution of recombination crossovers and the origin of haplotype blocks: the interplay of population history, recombination, and mutation. *Am. J. Hum. Genet.* **71** (5): 1227 – 34.
- WATKINS W. S., ZENGER R., O'BRIEN E., NYMAN D., ERIKSSON A. W., RENLUND M. & JORDE L. B. 1994. Linkage disequilibrium patterns vary with chromosomal location: A case study from the von Willebrand factor region. *Am. J. Hum. Genet.* **55**: 348 – 355.

Bibliografía

- WATTERSON G. A. y GUESS H. A. 1977. Is the most frequent allele the oldest?. *Theor. Popul. Biol.* **11**: 141 – 160.
- WEBER J. L. & MAY P. E. 1989. Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. *Am. J. Hum. Genet.* **44**: 388 -396.
- WEBER J. L., WONG C. 1993. Mutation of human short tandem repeats. *Hum. Mol. Genet.* **2** (8): 1123 – 8.
- WEBER J. L. 1990. Informativeness of human (dC-dA)_n'(dG-dT)_n polymorphisms. *Genomics* **7**: 524 – 530.
- WEI C.-C., CHIANG F.-T., LIN K.-S. & LIN L.-I. 1999. The Spectrum of Microsatellite Loci on Chromosomes 7 & 8 in Taiwan Aboriginal populations: a Comparative Population Genetic Study. *Hum. Genet.*, **104**: 333 – 340.
- WEIR B. S. 1979. Inferences about linkage disequilibrium. *Biometrics* **35**: 235 – 254.
- WEIR B. S. & COCKERHAM C. C. 1973. Mixed self and random mating at two loci. *Genet. Res.* **21**: 247 – 252.
- WEIR B. S. & COCKERHAM C. C. 1978. Testing hypothesis about linkage disequilibrium with multiple alleles. *Genetics* **88**: 633 – 642.
- WEIR B. S. & COCKERHAM C. C. 1984. Estimating F-Statistics for the Analysis of Population Structure. *Evolution* **38**: 1358 – 1370.
- WEIR B. S., ALLARD R. W. & KAHLER A. L. 1972. Analysis of complex allozyme polymorphism in a Barley Population. *Genetics* **72**: 505 – 523.
- WEISS K. M. & TERWILLIGER J. D. 2000. How many diseases does it take to map a gene with SNPs?. *Nat. Genet.* **26** (2): 151 – 7.
- WEISSENSTEINER T. & LANCHBURY J. S. 1996. Strategy for controlling preferential amplification and avoiding false negatives in PCR typing. *Biotechniques* **21** (6): 1102 – 8.
- WHITE J. K., AUERBACH W., DUYAO M. P., VONSATTEL J. P., GUSELLA J. F., JOYNER A. L., MACDONALD M. E. 1997. Huntingtin is required for neurogenesis and is not impaired by the Huntington's disease CAG expansion. *Nat. Genet.* **17** (4): 404 – 10.
- WIEHE T. 1998. The effect of selective sweeps on the variance of the allele distribution of a linked multiallele locus: Hitchhiking of microsatellites. *Theor. Pop. Biol.* **53**: 272 – 283.
- WIERDL M., DOMINSKA M., PETES T. D. 1997. Microsatellite instability in yeast: dependence on the length of the microsatellite. *Genetics* **146** (3): 769 – 79.
- WILLING M. C., DESCHENES S. P., SCOTT D. A., BYERS P. H., SLAYTON R. L., PITTS S. H., ARIKAT H., ROBERTS E. J. 1994. Osteogenesis imperfecta type I: molecular heterogeneity for COL1A1 null alleles of type I collagen. *Am. J. Hum. Genet.* **55**: 638 – 47.

- WILLING M. C., PRUCHNO C. J., ATKINSON M., BYERS P. H. 1992. Osteogenesis imperfecta type I is commonly due to a COL1A1 null allele of type I collagen. *Am. J. Hum. Genet.* **51**: 508 – 15.
- WILLING M. C., DESCHENES S. P., SLAYTON R. L., ROBERTS E. J. 1996. Premature chain termination is a unifying mechanism for COL1A1 null alleles in osteogenesis imperfecta type I cell strains. *Am. J. Hum. Genet.* **59** (4): 799 – 809.
- WILSON J. F. & GOLDSTEIN D. B. 2000. Consistent long-range linkage disequilibrium generated by admixture in a Bantu-Semitic hybrid population. *Am. J. Hum. Genet.* **67**(4): 926 – 35.
- WILSON J. F., WEALE M. E., SMITH A. C., GRATRIX F., FLETCHER B., THOMAS M. G., BRADMAN N., GOLDSTEIN D. B. 2001. Population genetic structure of variable drug response. *Nature Genetics* **29**: 265 – 269.
- WRIGHT A. F., CAROTHERS A. D. & PIRASTU M. 1999. Population choice in mapping genes for complex diseases. *Nat. Genet.* **23**: 397 – 404.
- WU Y., STULP R. P., ELFFERICH P., OSINGA J., BUYS C. H., HOFSTRA R. M. 1999. Improved mutation detection in GC-rich DNA fragments by combined DGGE and CDGE. *Nucleic Acids Res.* **27** (15): e9.
- WU Y., HAYES V. M., OSINGA J., MULDER I. M., LOOMAN M. W., BUYS C. H., HOFSTRA R. M. 1998. Improvement of fragment and primer selection for mutation detection by denaturing gradient gel electrophoresis. *Nucleic. Acids Res.* **26** (23): 5432 – 40.
- XIONG M. & JIN L. 1999. Comparison of the power and accuracy of biallelic and microsatellite markers in population-based gene-mapping methods. *Am. J. Hum. Genet.* **64**: 629 – 640.
- XU H. & FU Y.-X. 2004. Estimating effective population size or mutation rate with microsatellites. *Genetics* **166**: 555 – 563.
- XU X., PENG M., FANG Z. 2000. The direction of microsatellite mutations is dependent upon allele length. *Nat. Genet.* **24** (4): 396 – 9.
- YADOLLAHI-FARSANI M., KEFALAS P., SAXTY B. A. & McDERMOT J. 1999. Polymorphic forms of human ADP-ribosyltransferase-1. Differences in their catalytic activities revealed by labeling of membrane-associated substrates. *Eur. J. Biochem.* **262**: 342 – 348.
- YASUDA N. 1968. Estimation of the inbreeding coefficient from phenotype frequencies by a method of maximum likelihood scoring. *Biometrics* **24** (4): 915 – 935.
- YBALLE C. M., VU T. H. & HOFFMAN A. R. 1996. Imprinting and expresión of Insulin-Like Growth Factor-II and H19 in normal breast tissue and breast tumor. *J. Clin. End. Metab.* **81**: 1607 – 1612.
- YU A., ZHAO C., FAN Y., JANG W., MUNGALL A. J., DELOUKAS P., OLSEN A., DOGGETT N. A., GHEBRANIOUS N., BROMAN K. W., WEBER J. L. 2001. Comparison of human genetic and sequence-based physical maps. *Nature* **409** (6822): 951 – 3.

Bibliografía

- ZAPATA C. & ALVAREZ G. 1987. Gametic disequilibrium in populations of *Drosophila subobscura*: A review of experimental evidence. *Genét. Ibér* **39**: 593 – 616.
- ZAPATA C. & ALVAREZ G. 1992. The detection of gametic disequilibrium between allozyme loci in natural populations of *Drosophila*. *Evolution* **46**: 1900 – 1917.
- ZAPATA C. & ALVAREZ G. 1993. On the detection of nonrandom associations between DNA polymorphisms in natural populations of *Drosophila*. *Mol. Biol. Evol.* **10**: 823 – 841.
- ZAPATA C. & ÁLVAREZ G. 1997. On Fisher's exact test for detecting gametic disequilibrium between DNA polymorphisms. *Ann. Hum. Genet.* **61**: 71 – 77.
- ZAPATA C. & ÁLVAREZ G. 1997. Testing for homogeneity of gametic disequilibrium among populations. *Evolution* **51**: 606 – 607.
- ZAPATA C. & VISEDO G. 1995. Gametic disequilibrium and physical distance. *Am. J. Hum. Genet.* **57**: 190 – 191.
- ZAPATA C. 2000. The D' measure of overall gametic disequilibrium between pairs of multiallelic loci. *Evolution* **54**: 1809 – 1812.
- ZAPATA C., ÁLVAREZ G. & CAROLLO C. 1997. Approximate variance of the standardized measure of gametic disequilibrium D' . *Am. J. Hum. Genet.* **61**: 771 – 774.
- ZAPATA C., ÁLVAREZ G., RODRÍGUEZ-TRELLES F., MASIDE X. 2000. A long-term study on seasonal changes of gametic disequilibrium between allozymes and inversions in *Drosophila subobscura*. *Evolution* **54**: 1673 – 1679.
- ZAPATA C., CAROLLO C. & RODRÍGUEZ S. 2001a. Sampling variance and distribution of the D' measure of overall gametic disequilibrium between multiallelic loci. *Ann. Hum. Genet.* **65**: 395 – 406.
- ZAPATA C., NÚÑEZ C. & VELASCO T. 2002. Distribution of nonrandom associations between pairs of protein loci along the third chromosome of *Drosophila melanogaster*. *Genetics* **161**: 1539 – 1550.
- ZAPATA C., RODRÍGUEZ S., VISEDO G. & SACRISTÁN F. 2001b. Spectrum of nonrandom associations between microsatellite loci in human chromosome 11p15. *Genetics* **158**: 1235 – 1251.
- ZAPATA C., RODRÍGUEZ S., RODRÍGUEZ L., HERMIDA J. & SANDE E. Sample-size requirements of the International HapMap Project. (Enviado).
- ZAVATTARI P., DEIDDA E., WHALEN M., LAMPIS R., MULARGIA A., LODDO M., EAVES I., MASTIO G., TODD J. A., CUCCA F. 2000. Major factors influencing linkage disequilibrium by analysis of different chromosome regions in distinct populations: demography, chromosome recombination frequency and selection. *Hum. Mol. Genet.* **9** (20): 2947 – 57.
- ZERBA K. E., KESSLING A. M., DAVIGNON J. & SING C. F. 1991. Genetic structure and the search for genotype-phenotype relationships: An example from disequilibrium in the Apo B gene region. *Genetics* **129**: 525 – 533.

- ZHANG K., AKEY J. M., WANG N., XIONG M., CHAKRABORTY R., JIN L. 2003. Randomly distributed crossovers may generate block-like patterns of linkage disequilibrium: an act of genetic drift. *Hum. Genet.* **113** (1): 51 – 9.
- ZHANG W., COLLINS A., GIBSON J., TAPPER W. J., HUNT S., DELOUKAS P., BENTLEY D. R., MORTON N. E. 2004. Impact of population structure, effective bottleneck time, and allele frequency on linkage disequilibrium maps. *Proc. Natl. Acad. Sci. USA.* **101**: 18075 – 18080.
- ZHAO H., MCPEEK M. S., SPEED T. P. 1995. Statistical analysis of chromatid interference. *Genetics* **139**: 1057 – 65.
- ZHIVOTOVSKY L. A. 1999. A new genetic distance with application to constrained variation at microsatellite loci. *Mol. Biol. Evol.* **16** (4): 467 – 71.
- ZHIVOTOVSKY L. A., BENNETT L., BOWCOCK A. M., FELDMAN M. W. 2000. Human population expansion and microsatellite variation. *Mol. Biol. Evol.* **17** (5): 757 – 67.
- ZONDERVAN K. T. & CARDON L. R. 2004. The complex interplay among factors that influence allelic association. *Nat. Rev. Genet.* **5**: 89 – 100.

