



UNIVERSIDAD NACIONAL AUTÓNOMA
DE MÉXICO

FACULTAD DE CIENCIAS

EVOLUCIÓN DE MICROSATÉLITES DE CLOROPLASTO EN *Pinus pseudostrobus* y *Pinus montezumae*: EFECTO DEL LIGAMIENTO EN LA ESTIMACIÓN DE PARÁMETROS POBLACIONALES Y APLICACIONES EN GENÉTICA DE POBLACIONES

T E S I S

QUE PARA OBTENER EL TÍTULO DE:

B I Ó L O G A

P R E S E N T A:

VALERIA ALAVEZ GOMEZ



TUTOR: DR. DANIEL IGNACIO PIÑERO DALMAU

2008



Universidad Nacional
Autónoma de México

Dirección General de Bibliotecas de la UNAM

Biblioteca Central



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

RESUMEN

Los microsatélites se han convertido en importantes marcadores genéticos debido a su presencia y abundancia en los genomas de gran variedad de organismos y a su elevado polimorfismo. En particular los microsatélites pertenecientes al genoma de cloroplasto, al estar ligados, no recombinar y ser heredados uniparentalmente, representan una herramienta útil para realizar estudios en el ámbito de las poblaciones. No obstante, también se han reportado resultados contradictorios que presentan ciertas dificultades vinculadas directamente con su proceso evolutivo, como lo son una tasa de mutación baja (en el cloroplasto), la cual puede ser insuficiente para generar el polimorfismo necesario para realizar estudios en el ámbito de las poblaciones, o bien, la incidencia de homoplasia molecular de tamaño, fenómeno esperado a tasas de mutación elevadas. A raíz de lo anterior, este trabajo pretende exponer el potencial y los alcances que pueden tener los microsatélites de cloroplasto como marcadores genéticos tomando en consideración los inconvenientes planteados. Para ello se realizaron simulaciones tomando como referencia secuencias de microsatélites encontrados en *Pinus pseudostrobus* Lamb y *P. montezumae* Lindl. e incorporando el modelo de mutación por pasos propuesto por Ohta y Kimura (1978).

Los resultados obtenidos señalan que la variación genética esperada para las tasas de mutación reportadas en el cloroplasto es adecuada para realizar estudios en el campo de la genética de poblaciones. Respecto a la homoplasia molecular se reconocieron tres aproximaciones de la misma: la homoplasia de tamaño (*SH*), la homoplasia molecularmente accesible (*MASH*) y la homoplasia estructuralmente accesible (*SASH*). *MASH* es accesible mediante técnicas de laboratorio y tiene la capacidad de revelar una importante proporción de *SH* por lo que constituye una solución importante a los problemas que acarrea la presencia de esta última. *SASH* es una medida de homoplasia accesible únicamente por simulaciones computacionales. La proporción de *SASH* al no poder ser descubierta como *MASH*, persiste como un ‘ruido’ inherente a los datos, no obstante decrece de manera importante al trabajar con la mayor cantidad de loci posible.

Así pues, debido al ligamiento de loci en el cloroplasto, el uso de los microsatélites es viable considerando el empleo de más de 2 loci para obtener una buena cantidad de

información en términos de variación genética y reducir lo más posible la homoplasia molecular.

Por otra parte, se construyeron redes de haplotipos a partir de muestras simuladas de microsatélites y se estudiaron dos aspectos fundamentalmente: a) la proporción con que el haplotipo más ancestral es también el más abundante y/o el más conectado, y b) el efecto que *SASH* tiene sobre la topología de las redes de haplotipos. Se encontró que el haplotipo más ancestral coincide sólo ocasionalmente con ser el más abundante y/o el más conectado cuando los valores de θ multilocus ($\theta = 2N_e\mu$) están por arriba de 5. *SASH* por su parte afecta a las redes simplificando su estructura al reducir el número de nodos y colapsar ciertas ramas.

ÍNDICE

I.	Introducción	6
	1.1 Microsatélites: Estructura y distribución	6
	1.2 Mecanismos de mutación	8
	1.3 Modelos de mutación	10
	1.4 Homoplasia molecular de tamaño	12
	1.5 Simulaciones y el coalescente	14
	1.6 Microsatélites y genealogías	16
II.	Objetivos	19
III.	Métodos	20
	3.1 Simulaciones	20
	3.1.1 Generación de genealogías	20
	3.1.2 Incorporación de mutaciones y configuraciones de muestra	20
	3.1.3 Microsatélites	21
	3.1.4 Parámetros iniciales	21
	3.1.5 Parámetros calculados	22
	3.2 Redes de haplotipos	22
	3.2.1 Medidas aplicadas a redes	23
IV.	Resultados	25
	4.1 Subconjuntos de la muestra: electromorfos, variantes multilocus y haplotipos	25

4.2 Heterocigosis	26
4.3 Número de variantes y haplotipos	29
4.4 Homoplasia	32
4.4.1 Homoplasia calculada a partir de la heterocigosis	32
4.4.2 Homoplasia basada en el coalescente	34
4.4.3 Homoplasia molecularmente accesible	36
4.5 Homoplasia y heterocigosis	39
4.6 Homoplasia y tamaño efectivo de las poblaciones	40
4.7 Homoplasia y θ	42
4.8 Haplotipos ancestrales y genealogías	44
4.8.1 Conectividad y abundancia en redes	46
4.8.2 Efecto de la homoplasia sobre la estructura de las redes	48
V. Discusión	52
5.1 Cálculos respecto a los subconjuntos de la muestra	52
5.2 Homoplasia molecularmente accesible	52
5.3 Homoplasia respecto a θ	54
5.4 Modelos de mutación	55
5.5 Genealogías de haplotipos	57
5.5.1 Relación entre el ancestro y la abundancia	57
5.5.2 Topología de redes y homoplasia	59
VI. Conclusiones y perspectivas	62
VII. Apéndice 1: Redes	64
VIII. Glosario	68
IX. Referencias	71

I. INTRODUCCIÓN

1.1 Microsatélites: estructura y distribución

La estructura del genoma de los organismos eucariontes se caracteriza por contener una gran proporción del llamado ADN repetitivo. De una manera general, el genoma de todos los organismos presenta regiones con secuencias únicas y regiones con secuencias redundantes, estas últimas correspondientes usualmente, mas no exclusivamente, a ADN no codificante. Aunque los genomas procariontes poseen estos elementos repetitivos, es en los genomas de organismos eucariontes donde están representados de manera sobresaliente (Griffiths *et al.*, 1999).

El ADN repetitivo está caracterizado por múltiples copias de distintos tipos de unidades (pares de bases sencillas, secuencias de nucleótidos, transposones, entre otros) que se distribuyen de manera dispersa o agrupada a lo largo del genoma. Entre las formas dispersas destacan los transposones y los distintos tipos de retroelementos: retrotransposones, elementos nucleares dispersos largos (*long interspersed nuclear elements*, LINE's) y cortos (*short interspersed nuclear elements*, SINE's), por mencionar algunos (Brown, 2002).

Entre las formas agrupadas, también denominadas ADN satelital, se encuentran los satélites, los minisatélites y los microsatélites (Brown, 2002), los cuales se distinguen entre sí fundamentalmente por su tamaño. El término ADN satelital surgió entre 1960 y 1970, a partir de la detección de una fracción de ADN que mostró una densidad de flotación muy distinta al realizar una centrifugación en un gradiente de densidad. Este tipo de ADN corresponde propiamente a los fragmentos ahora reconocidos como 'satélites' los cuales son arreglos donde miles de repeticiones de 2 a cientos de pares de bases constituyen un gran porcentaje de todo el genoma y ocurren en arreglos individuales que se extienden hasta 5 megabases. Se ubican preferentemente en la heterocromatina de los centrómeros

(Ellegren, 2004). Más pequeños que estos últimos, los minisatélites pueden presentarse en cientos o incluso miles de loci diferentes por genoma y la unidad que se repite puede estar constituida por una secuencia larga (> 10 pares de bases) lo cual resulta en bloques de entre 0.5 a 30 kilobases (Armour *et al.*, 1999).

Los microsatélites se definen como conjuntos de uno a ocho pares de bases contiguas repetidas de 10 a 100 veces, por lo que corresponden a iteraciones agrupadas de unidades que conforman arreglos cortos que se prolongan en promedio de 20 a 100 pares de bases por locus (Armour *et al.*, 1999). Por ejemplo, en humanos es común encontrar una gran proporción de microsatélites dinucleótidos del tipo (CA)_n o bien:



No obstante, existe una gran diversidad de configuraciones de microsatélites, desde mononucleótidos simples, como:



o bien, di-, tri- y tetranucleótidos, o hasta penta- o hexanucleótidos (Ellegren, 2004). Cuando se intercalan otros nucleótidos entre las distintas repeticiones de denominan microsatélites interrumpidos:



y cuando se encuentran dos tipos de repeticiones contiguas, se denominan compuestos:



En general, se ha observado que los microsatélites se presentan ubicuamente en todos los genomas eucariontes, pero su número y proporción respecto al genoma en que se encuentran varía significativamente entre especies y entre niveles taxonómicos mayores (Ellegren, 2004). Por ejemplo, la densidad microsatelital tiende a correlacionarse positivamente con el tamaño del genoma de la mayoría de los organismos eucariontes,

salvo en las plantas donde dicha correlación es negativa según estudios realizados por Morgante y colaboradores (2002), quienes compararon la frecuencia de microsatélites respecto al tamaño del genoma de distintos grupos de plantas. Dicha evidencia indica que la mayoría de los microsatélites se encuentran en regiones ancestrales que preceden la reciente expansión genómica (atribuida a otras regiones repetitivas como los retrotransposones) en muchos grupos de plantas (Morgante *et al.*, 2002).

Dos atributos convierten a los microsatélites en una herramienta muy útil para desarrollar estudios en el campo de la genética de poblaciones, donde el reconocimiento de alelos en determinados loci y la cuantificación de sus frecuencias permite hacer inferencias históricas y demográficas importantes: uno es su alta variabilidad y el segundo es su densa distribución en los genomas eucariontes. El polimorfismo en los microsatélites está dado por la variación de su tamaño, es decir, del número de unidades que los estructuran.

La mayoría de los microsatélites que se emplean como marcadores en estudios de las poblaciones se encuentran asociados a ADN no codificante, en las secuencias intergenéticas o en intrones, por ello generalmente se supone que no existen presiones selectivas sobre ellos, y por tanto, que evolucionan neutralmente. No obstante, su ubicación en el genoma no está restringida a la porción que no codifica. Otro grupo de microsatélites se encuentra asociado a regiones promotoras, sitios de unión a proteínas o cerca de estos e incluso a ADN codificante (Ishii *et al.*, 2001). Estos últimos, están altamente conservados ya que alteraciones en su tamaño pueden conducir a patologías o al mal funcionamiento de ciertas proteínas (Kashi y Soller, 1999; Rubinsztein, 1999).

Por otra parte, los microsatélites no sólo se encuentran inmersos en el genoma nuclear, sino también en los genomas de mitocondrias y de cloroplastos. Los microsatélites de cloroplasto han sido empleados exitosamente en estudios de muchos grupos de plantas (Doyle *et al.*, 1998; Ishii *et al.*, 2001; Ribeiro *et al.*, 2001; Provan *et al.*, 2001; Mogg *et al.*, 2002). Es importante recalcar que estos presentan un par de características que los distinguen de los microsatélites nucleares: los cloroplastos se heredan uniparentalmente, lo cual implica el seguimiento de linajes de un solo sexo y por otra parte no recombinan por lo que todos sus loci están ligados (Provan *et al.*, 2001; Navascués y Emerson, 2005), por lo tanto, su evolución es sencilla de modelar bajo un enfoque coalescente.

1.2 Mecanismos de mutación

Siendo los microsatélites iteraciones de unidades compuestas por nucleótidos, las mutaciones de inserción/delección alteran su tamaño al variar el número de unidades que los componen. En otras palabras, alargan o acortan el locus a medida que se añaden o sustraen unidades al momento de cada evento de mutación. Cada alelo microsatelital se distingue pues por su longitud.

Los cambios de longitud en los microsatélites se atribuyen generalmente al desplazamiento y a la desalineación de las hebras de ADN durante la duplicación (*replication slippage*). Mientras ocurre la síntesis, la hebra recién sintetizada puede disociarse de la hebra líder o templado y reasociarse de manera desalineada, probablemente a causa de la formación de asas en alguna de las dos hebras. En consecuencia, como producto de la polimerización, resulta una hebra hija poseedora de un microsatélite alargado o acortado en longitud por una unidad (Levinson, 1987). Si el asa ocurre en la hebra templado, el microsatélite resultante será más corto, e inversamente será más largo si el asa ocurre en la hebra naciente (Fig. 1.1).

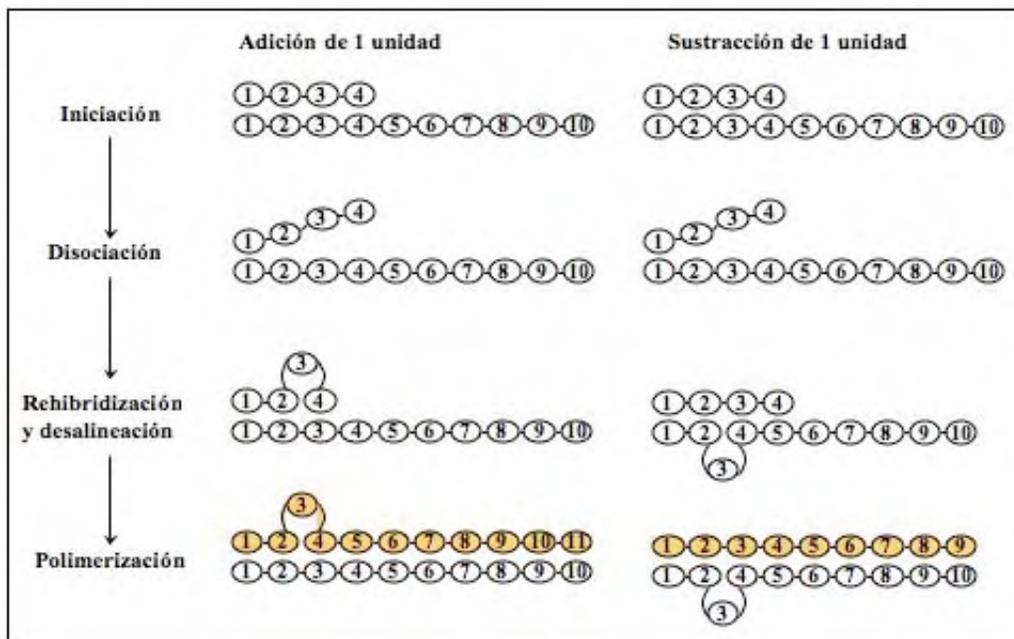


Figura 1.1 Desplazamiento y desalineación de las hebras de ADN (*Replication slippage*). Las hebras resultantes (en amarillo) son de distinto tamaño que el templado.

No obstante, un sistema molecular de reparación de incompatibilidad (*mismatch repair system*) se contrapone al proceso anterior al detectar regiones donde las hebras no se acoplan correctamente, lo que a su vez facilita la remoción y desintegración de las hebras mal sintetizadas. Un mecanismo como éste reduce la incidencia de mutaciones ocasionadas por el desalineamiento y se traduce en la reducción de la tasa de mutación, la cual es, consecuentemente, producto de dos mecanismos: el desplazamiento y desalineación de las hebras de ADN y la eficiencia del mecanismo de reparación (Goldstein y Schlötterer, 1999).

Estas mutaciones ocurren a tasas de varios órdenes de magnitud más que las mutaciones puntuales (Provan *et al.*, 1999; Hancock, 1999). Las tasas de sustitución comúnmente oscilan entre 10^{-10} y 10^{-8} eventos por locus por generación tanto en el núcleo como en la mitocondria y el cloroplasto (Provan *et al.*, 1999), mientras que se calcula que los microsatélites nucleares poseen tasas de alrededor de 10^{-5} a 10^{-4} en la levadura (*Saccharomyces cerevisiae*), 10^{-3} en el humano, de 10^{-4} a 10^{-3} en el ratón (*Mus musculus*) y 6×10^{-6} en la mosca de la fruta (*Drosophila melanogaster*) (Hancock, 1999). Por otra parte, para el caso del genoma del cloroplasto se ha reportado que las tasas de mutación son más lentas que en el núcleo e incluso se ha discutido si son suficientes para generar polimorfismo (Provan *et al.*, 1999; Provan *et al.*, 2001), no obstante, se han documentado tasas de sustitución de 1×10^{-9} a 3×10^{-9} (Provan *et al.*, 1999) en tanto que los escasos cálculos de las tasas de mutación para microsatélites oscilan entre 3.7×10^{-6} a 10^{-3} mutaciones por locus por generación, reportados para *Arabidopsis thaliana* (Jakobsson *et al.*, 2007) y para *Pinus contorta* (Marshall *et al.*, 2002) respectivamente.

Por otra parte, se ha sugerido que un mecanismo de recombinación desigual puede ser responsable de la adición y sustracción de unidades microsatelitales provocando las variaciones de longitud, no obstante, poca evidencia sustenta esta noción (Ellegren, 2004). En el genoma nuclear del humano, la distribución de ciertos microsatélites está asociada a sitios recombinantes, sin embargo, esta relación parece, por un lado, revelar que ciertas secuencias repetitivas están involucradas en el proceso de recombinación y por otro, negar que son una consecuencia de éste, ya que existe poca correlación entre la tasa de recombinación y la densidad y mutabilidad de los microsatélites (Trecó *et al.*, 1986;

Majewski y Ott, 2000). Se suma a estas observaciones evidencia de que no existe diferencia en las tasas y las formas de mutación entre los microsatélites de cromosomas autosómicos y los ligados al cromosoma Y, el cual no lleva a cabo recombinación meiótica (Kayser *et al.*, 2000; Heyer *et al.*, 1997).

1.3 Modelos de mutación

Al estudiar los datos empíricos obtenidos mediante el uso de microsatélites para calcular diversos parámetros relativos a las poblaciones y distancias genéticas, es necesario aplicar un modelo teórico de evolución de microsatélites que se ajuste a las distribuciones de frecuencias alélicas encontradas en las poblaciones naturales.

El modelo fundamental que describe el proceso evolutivo de los microsatélites, es el modelo de mutación por pasos (Ohta y Kimura, 1978) o stepwise mutation model (SMM, por sus siglas en inglés). Éste propone que las mutaciones modifican la longitud del microsatélite al añadir o remover, simétricamente, una de sus unidades a una tasa de mutación fija, independientemente de la longitud inicial del microsatélite, no obstante, ciertos estudios han revelado atributos relativos a la forma de mutar que son incompatibles con el modelo de mutación por pasos (Ellegren, 2004). Por ejemplo, los microsatélites largos tienen mayor probabilidad de desplazarse y desalinearse al comienzo de la replicación de ADN ya que presentan mayor superficie de realineación que aquellos más cortos y por tanto se esperaría que la tasa de mutación no fuera fija, sino que aumentara en función de la longitud del locus. Wierdl *et al.*, (1997) apoyan experimentalmente esta noción, al igual que Petes *et al.*, (1997) quienes encontraron que los microsatélites interrumpidos limitan el intervalo de realineación, ya que se rompen en fragmentos más cortos, y como consecuencia el desplazamiento de hebras ocurre con menor frecuencia, lo cual a su vez se traduce en una tasa de mutación menor. Por su parte, Jakobsson y colaboradores (2007) encontraron una relación lineal entre la tasa de mutación y la longitud de microsatélites de cloroplasto en *Arabidopsis thaliana* (L.) Heynh. cuando esta última rebasa las 7 unidades. Resultados similares han sido reportados por Rose y Falush (1998) en microsatélites de levadura, por Sibly *et al.*, (2003) en humano y Goldstein y Clark (1995) en *Drosophila melanogaster*.

Por otra parte, también se ha discutido que los microsatélites parecen alcanzar un tamaño superior límite (Garza *et al.*, 1995), lo cual es incompatible con el modelo de mutación por pasos, ya que de acuerdo a éste se esperaría encontrar alelos muy largos, muy cortos o incluso, la desaparición de los mismos. Para explicar dicho tamaño máximo, se ha sugerido que existe una tendencia de mutación de los alelos de mayor tamaño a perder unidades (Calabrese *et al.*, 2001) o bien que se presenta un equilibrio entre las mutaciones debidas al desplazamiento de hebras y las mutaciones de sustitución, de manera que estas últimas fragmentan a los microsatélites de elevada longitud (Kruglyak *et al.*, 1998).

Di Rienzo *et al.* (1994), discuten el modelo de mutación por pasos respecto al número de unidades que se adicionan o sustraen por cada evento de mutación. Proponen un modelo de dos fases (*two phase mutation model*) donde la mayoría de las mutaciones suceden de acuerdo al modelo por pasos, sin embargo, incorporan cambios no frecuentes de más de una unidad.

En un estudio comparativo se probó el ajuste de diversos modelos evolutivos mediante el uso de datos genómicos de loci microsatelitales homólogos en humanos y en chimpancés (Sainudiin *et al.*, 2004). Un modelo de mutación por pasos modificado resultó ser el más adecuado. Éste se caracteriza por proponer la sustracción o adición de únicamente una unidad por mutación, a una tasa con una tendencia lineal, proporcional al tamaño del microsatélite. La variación de tamaño de más de una unidad por mutación falló en mostrar un mejor ajuste.

Dado que los microsatélites han representado una herramienta sumamente popular en los últimos años, particularmente en los campos de la genética de poblaciones, la filogeografía y la demografía, se ha generado inquietud respecto a la validez de su uso como marcadores para desarrollar estudios en los ámbitos mencionados, ya que las inferencias generadas pueden ser sensibles al modelo supuesto de evolución de microsatélites ya que, como se expone arriba, no hay un único modelo que describa todos los comportamientos relacionados a la forma de mutar observados empíricamente. Además, se suma al problema de la selección de modelos, el fenómeno conocido como homoplasia molecular de tamaño (Estoup *et al.*, 2002) el cual merece la pena tratar con mayor detalle.

1.4 Homoplasia molecular de tamaño

Como se ha mencionado, el mecanismo de mutación que actúa sobre los microsatélites resulta en la variación de su tamaño. Un par de microsatélites del mismo tamaño, es decir idénticos por estado, pueden compartir esta homología estructural, a través de eventos de mutación independientes, sin necesariamente provenir del mismo ancestro. Por ejemplo, la figura 1.2 muestra una genealogía sencilla de tres generaciones de microsatélites del presente (t_0) hacia atrás en el tiempo (t_1 y t_2), en la que dos eventos de mutación independientes se presentan produciendo alelos de igual tamaño (b y c) pero originados a partir de mutaciones distintas. Este fenómeno en el que los microsatélites son idénticos por estado, pero difieren por descendencia es conocido como homoplasia de tamaño (*size homoplasy* o *SH*; Estoup *et al.*, 2002).

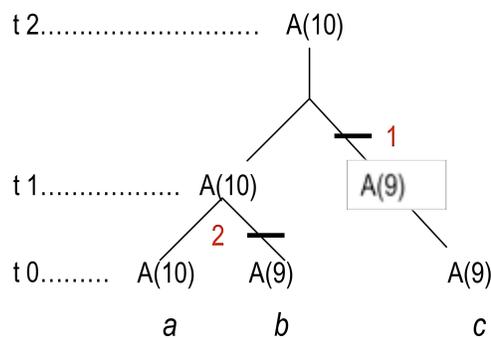


Figura 1.2 Genealogía de microsatélites con mutaciones superpuestas, mostrando tres generaciones, del presente (t_0) hacia atrás en el tiempo (t_1 y t_2). Los alelos b y c son homoplásicos al presentar el mismo estado, $A(9)$, sin compartir la misma historia mutacional.

Una primera consecuencia de la homoplasia es la subvaloración de la diversidad genética. Dos marcadores que convergen en tamaño pero que no provienen del mismo linaje son diferentes ya que al ser producto de mutaciones independientes que ocurrieron en tiempos distintos, poseen una historia evolutiva propia, por lo tanto son haplotipos distintos. La homoplasia genera variantes de tamaño idéntico (electromorfos) que pueden contener haplotipos que resultan de distintos linajes, de manera que habrá menos variantes que haplotipos. Es evidente, en consecuencia, que la homoplasia puede causar alteraciones en estudios filogenéticos y genealógicos al oscurecer las relaciones de ancestría y en análisis de genética de poblaciones, al afectar de manera directa medidas fundamentales como lo son la diversidad, el flujo y las distancias genéticas.

Por un lado, un recurso para reducir los efectos de la homoplasia de tamaño es incrementar el número de loci de microsatélites ligados. Esto aumenta la resolución en el proceso de identificación de haplotipos ya que aunque uno de los loci no presente variación, es decir sea monomórfico, existe la posibilidad de que algún locus vecino sea polimórfico (Navascués y Emerson, 2005). Por otro lado, cuando se estudia un microsatélite compuesto, el conocimiento de la secuencia de cada uno de los electromorfos, mediante herramientas como la secuenciación o el polimorfismo de conformación de cadena individual de ADN (*single-strand conformation polymorphism* o SSCP), puede desentrañar diferencias básicas entre éstas. Por ejemplo, dos variantes idénticas por tamaño como serían A(8)-G(9) y A(7)-G(10) migrarán equivalentemente en una electroforesis, sin embargo, mediante el conocimiento de su secuencia se puede determinar que se trata de dos haplotipos sustancialmente diferentes. Este atributo se reconoce como homoplasia molecularmente accesible (*molecularly accessible size homoplasia* o *MASH*; Estoup *et al.*, 2002) y representa una fracción de la homoplasia por tamaño (*SH*).

Se discute aún en qué medida la homoplasia influye en los análisis en el ámbito de las poblaciones. Ciertos estudios reportan importantes alteraciones a los cálculos de diversidad genética, por ejemplo, reducción del número observado de alelos por población y disminución de la heterocigosis (Angers *et al.*, 2000; van Oppen *et al.*, 2000), o bien decremento de la precisión en la prueba de neutralidad para expansión de las poblaciones (F_S) y en el cálculo de parámetros demográficos proporcional con la magnitud de homoplasia presente en los datos (Navascués *et al.*, 2006). Por otra parte, otros estudios revelan que la homoplasia tuvo poco impacto en las medidas de diferenciación y estructuración genética, como lo son los índices de fijación R_{ST} y F_{ST} (Angers *et al.*, 2000, Adams *et al.*, 2004).

La incorporación de estudios mediante simulaciones ha sido un acercamiento útil al problema ya que ha permitido la valoración de la homoplasia en función de distintas combinaciones de parámetros, no obstante, también bajo este enfoque se han obtenido resultados dispares. Por un lado, se presenta un panorama desalentador para continuar con el uso de los microsatélites y se recomienda emplear la mayor cantidad de loci posibles para contrarrestar el efecto de la homoplasia sobre la diversidad genética (Navascués y

Emerson, 2005), y por otro, se dice que la homoplasia se mantiene en proporciones bajas para ciertos grupos de parámetros (Estoup *et al.*, 2002).

1.5 Simulaciones y el coalescente

Es evidente la necesidad de estudiar con mayor profundidad los atributos relativos a la mutación y a la evolución de los microsatélites y su implicación, en este caso, en el cálculo de medidas poblacionales fundamentales, ya que son éstas las que dan evidencia de los procesos que subyacen a la evolución de las poblaciones. Con este fin, un estudio basado en simulaciones apoyado en la teoría coalescente puede despejar algunas de las inquietudes arriba expuestas, con la ventaja que ofrece la posibilidad de controlar las variables para entender la influencia de distintos parámetros como lo son el tamaño efectivo de la población, la tasa de mutación, el tamaño de muestra y la estructura del microsatélite (sencillo o compuesto).

La teoría de coalescencia aporta una base sobre la cual es posible modelar distintos escenarios que afectan la variación genética como la demografía o las fuerzas evolutivas (deriva génica, selección natural, mutación y migración) e incorporar algoritmos computacionales para generar gran cantidad de datos simulados. El coalescente se fundamenta en que todos los estados alélicos actuales de cierto locus en una población están determinados por su historia mutacional y genealógica de manera que conociendo una muestra de alelos en el presente se puede modelar su historia pasada yendo hacia atrás en el tiempo. En otras palabras, cada uno de los alelos en una muestra desciende de un alelo en la generación previa, los cuales a su vez descienden de un alelo en una generación anterior y así sucesivamente. El coalescente calcula generación a generación la probabilidad y el tiempo en que dos alelos encuentran su ancestría común hasta que finalmente se llega al ancestro común de toda la muestra. El producto de este modelo es la generación de un linaje que refleja las relaciones genealógicas entre los haplotipos, de manera que una vez establecida una genealogía se puede superponer a ésta un conjunto de mutaciones y a su vez simular gran cantidad de datos (Nordborg, 2001; Hein *et al.*, 2005).

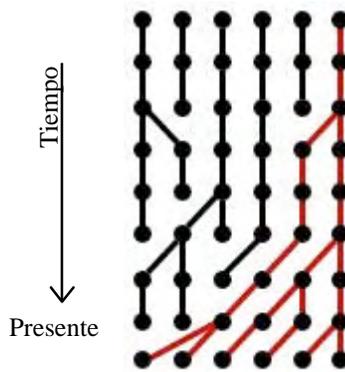


Figura 1.3 Diagrama de un proceso coalescente. En rojo se muestra la coalescencia de los linajes hasta encontrar un ancestro común.

Si se trabaja con alelos o variantes selectivamente neutras, el éxito reproductor no se ve afectado por los diferentes estados alélicos, de manera que es posible separar el proceso genealógico del proceso de mutación. Considerando, por ejemplo, el modelo básico de reproducción haploide propuesto por Wright y Fisher¹, el número de descendientes que puede dejar un gen particular es una variable estocástica cuya distribución probabilista es de carácter binomial en relación con los parámetros $2N$ (número de individuos presentes en cada generación) y $1/2N$ (probabilidad con que, viendo hacia atrás en el tiempo, un descendiente tiene un cierto progenitor en la generación previa). Las relaciones genealógicas determinadas por este modelo se pueden representar gráficamente (Fig. 1.3). Yendo hacia atrás en el tiempo, los linajes coalescen cada vez que dos o más individuos (genes o secuencias) encuentran el mismo progenitor en la generación anterior. Siguiendo este proceso, el número de linajes decrece y, con el paso de las generaciones, se reduce a un único linaje correspondiente al ancestro común más reciente (*most recent common ancestor* o MRCA, por sus siglas en inglés). Nada de lo anterior se ve afectado por las diferencias genéticas entre cada individuo ya que son neutrales (Nordborg, 2001).

Una vez modelada la genealogía, se puede asignar una identidad, es decir un tipo alélico, a cada individuo de la muestra. Para ello, se asigna un estado al ancestro común más reciente y se incorporan mutaciones en las ramas de la genealogía construida de acuerdo a un modelo de mutación adecuado. Siguiendo el proceso, ahora de atrás hacia

¹ El modelo de reproducción Wright-Fisher supone que las generaciones son discretas y no se superponen. El tamaño de las poblaciones se mantiene constante y todos los individuos (genes, secuencias) tienen la misma eficacia, por lo que no hay selección. La población no posee estructura geográfica ni social. Los genes o secuencias de la población no se recombinan (Hein *et al.*, 2005).

adelante en el tiempo, los descendientes en cada generación heredan el estado alélico de su progenitor en la generación previa, siempre y cuando no haya ocurrido una mutación en la transición de una generación a otra (Nordborg, 2001).

1.6 Microsatélites y genealogías

Los microsatélites polimórficos dentro de una especie han resultado también muy útiles para establecer relaciones genealógicas entre haplotipos o alelos. Dichas relaciones a su vez se emplean como un medio para analizar y poner a prueba una variedad de hipótesis evolutivas vinculadas a la teoría de genética de poblaciones (Posada y Crandall, 2001).

Como se abordó anteriormente, el proceso de duplicación de ADN es el responsable, por un lado, de generar copias idénticas de haplotipos de microsatélites dentro de una población, y por otro, cuando las hebras se desalinean, de originar nuevos haplotipos. Cuando se produce un nuevo alelo a partir de una de las copias de un tipo ancestral, el resto de éstas prevalece en la población y, en consecuencia, una muestra contendrá tanto haplotipos derivados como ancestrales. Debido a que los haplotipos ancestrales han permanecido en la población durante más tiempo es más probable que éstos hayan dado origen a nuevos haplotipos y, por tanto, se espera que la relación entre los antepasados y los descendientes sea multifurcada, de manera que una representación reticulada, donde el mayor número de conexiones corresponda al alelo o a los alelos más viejos de la población, es la más adecuada para exponer las relaciones genealógicas entre los haplotipos (Posada y Crandall, 2001).

Derivado de lo anterior y con base en predicciones de la teoría de genética de poblaciones, se han planteado una serie de generalizaciones en relación a que la topología de una red de haplotipos debería revelar que:

- Existe una relación directa entre la frecuencia y la edad de los haplotipos. Específicamente, la probabilidad de que un alelo representado n_i veces en una muestra de tamaño n sea el más ancestral es igual a n_i/n y el orden esperado de los alelos por edad es igual al orden de los alelos por frecuencia. Por lo tanto, los

haplotipos más frecuentes probablemente han estado presentes más tiempo en la población.

- Los alelos más antiguos (aquellos con alta frecuencia dentro de la población) poseen una mayor probabilidad de ser haplotipos interiores en una red (aquellos haplotipos con más de una conexión mutacional).
- En promedio, los alelos más antiguos tendrán una mayor distribución geográfica.
- Los haplotipos con mayor frecuencia tenderán a presentar mayor cantidad de conexiones.
- Los nuevos mutantes, en particular aquellos representados en la muestra por un solo individuo (*singletons*) se conectan con mayor probabilidad a haplotipos comunes (más conectados o con mayor frecuencia) que a otros haplotipos derivados.
- Es más probable que los haplotipos representados en la muestra por un solo individuo estén más relacionados a haplotipos de la misma población, que a haplotipos de otras poblaciones (Posada y Crandall, 2001).

Así pues, las redes de haplotipos constituyen una herramienta importante en los estudios genéticos de las poblaciones, cuyas inferencias evidentemente dependen de la precisión con las que las redes puedan ser construidas. Las simulaciones realizadas en este estudio abren la posibilidad de analizar con mayor profundidad las generalizaciones expuestas arriba, específicamente para el caso de los microsatélites donde la homoplusia representa un primer factor que sin duda altera la estructura de las redes, en principio porque implica un déficit de variación genética al subvalorar el número de haplotipos.

II. OBJETIVOS

El objetivo general de este trabajo es estudiar la evolución de los microsatélites de cloroplasto, mediante simulaciones computacionales, bajo un modelo de mutación por pasos simétrico (Ohta y Kimura, 1978), con el fin de señalar tanto las aptitudes como los inconvenientes de éstos como marcadores moleculares para el cálculo de parámetros de las poblaciones, primordialmente cuando los loci se encuentran ligados.

Los objetivos particulares son los siguientes:

- i) Estudiar la incidencia de la homoplasia molecular, distinguir sus distintos tipos y explorar la influencia de cada uno de estos, cuantificando la proporción en que se presentan, sobre parámetros de variación genética como lo son la heterocigosis o bien, el número de haplotipos y/o variantes.
- ii) Enfatizar el efecto que el ligamiento de loci confiere a los distintos parámetros, particularmente en la homoplasia molecular.
- iii) Explorar la influencia de la homoplasia sobre la estructura de redes de haplotipos.
- iv) Estudiar las relaciones genealógicas entre haplotipos y/o variantes, analizando particularmente el vínculo entre la frecuencia y la posición ancestral de estos, empleando muestras generadas en computadora.

III. MÉTODOS

3.1 Simulaciones

Las simulaciones se llevaron a cabo en el programa *gye-coalescence* (Winberger y Piñero, 2007) diseñado para simular la evolución de secuencias de microsatélites. El programa corre en cualquier plataforma que posea ambiente *Java 5 JRE (Java Runtime Environment)*.

3.1.1 Generación de genealogías. Las simulaciones para generar cada genealogía se realizaron siguiendo un razonamiento coalescente y aplicando un algoritmo derivado de éste de acuerdo a Hein *et al.* (2005). El algoritmo plantea en primera instancia comenzar con una muestra de alelos de tamaño k , tomados de una población con determinado tamaño efectivo de la población N_e . Los alelos considerados corresponden a microsatélites de cloroplasto encontrados en *Pinus pseudostrobus* y *P. montezumae*. A partir de esto se simula el tiempo de espera T_k^c , distribuido exponencialmente para cada evento coalescente de manera que $T_k^c \sim Exp((\binom{k}{2}))$. En otras palabras, se elige al azar un par de linajes (i, j) , donde $1 \leq i < j \leq k$ dentro de los posibles pares $\binom{k}{2}$. A continuación se encuentra el ancestro de i y j por lo que después de cada evento coalescente la muestra k se reduce en 1 o bien $k \rightarrow k - 1$. Este proceso se repitió hasta que sólo un linaje quedó presente en la muestra, esto es $k = 1$. El resultado es la generación de una genealogía de la muestra k de linajes.

3.1.2 Incorporación de mutaciones y configuraciones de muestra. Las mutaciones fueron añadidas a la genealogía generada previamente, a partir de la identidad asignada al ancestro común más reciente, la cual en cada caso se estableció como la variante alélica más frecuente de las encontradas por Delgado *et al.* (2007) en *P. pseudostrobus* y *P. montezumae*.

La incorporación de mutaciones se simuló de acuerdo a una distribución de Poisson, ya que ésta expresa la probabilidad de que un número de eventos (las mutaciones en este caso) ocurran a una tasa constante en un periodo temporal discreto provisto por el coalescente.

El modelo bajo el cual se incorporaron las mutaciones a la genealogía corresponde al modelo simétrico de mutación por pasos (*stepwise mutation model*, SMM) propuesto por Ohta y Kimura (1973), el cual supone la pérdida o ganancia, con igual probabilidad, de una unidad de microsatélite, a una tasa de mutación fija.

Cada mutación se registró numéricamente de acuerdo al momento en que surgió, siendo la mutación número 1 la más ancestral. Así, cada haplotipo se identificó por el conjunto de mutaciones que le dio origen. Como consecuencia, los alelos de los datos generados se agruparon de acuerdo a tres perspectivas distintas: a) considerando su tamaño, como se haría tras una electroforesis en el laboratorio; b) considerando su identidad aportada por las mutaciones conocidas únicamente a través de las simulaciones; y c) para el caso de los microsatélites compuestos o ligados, contemplando la secuencia de cada variante, la cual revela diferencias estructurales de cada locus. Por lo anterior, se definieron tres tipos de muestra: variantes por tamaño, haplotipos por identidad y variantes multilocus (por secuencia), respectivamente.

3.1.3 Microsatélites. Los microsatélites que se emplearon en la simulaciones corresponden a electromorfos polimórficos del genoma de cloroplasto de *Pinus pseudostrobus* y *Pinus montezumae* amplificados por Delgado *et al.* (2007) empleando oligonucleótidos derivados de la secuencia del genoma de cloroplasto de *Pinus thunbergii* (Pt1254, Pt15169, Pt30204, Pt71936, Pt9383 y Pt87268). Muestras de cada uno de estos loci fueron secuenciadas (Delgado *et al.*, 2007) y en cada simulación se asumió a la variante más frecuente de cada microsatélite como el ancestro común más reciente (MRCA; *most recent common ancestor*), ignorando las secuencias flanqueantes.

Cuadro 3.1 Secuencias más frecuentes de los microsatélites amplificados para *P. pseudostrobus* y *P. montezumae* por Delgado *et al.* (2007), ignorando las secuencias flanqueantes

Marcador	Secuencia	Frecuencia (n = 303)
Pt1254	T(11)	0.8416
Pt15169	T(8)	0.7119
Pt30204	A(8)-G(9)	0.6898
Pt71936	T(15)	0.46
Pt9383	T(7)-A(9)	0.6667
Pt87268	T(11)	0.4319

3.1.4 *Parámetros iniciales.* Los parámetros que es necesario especificar al momento de correr cada simulación son: el tamaño efectivo de la población N_e , la tasa de mutación μ , el tamaño de la muestra k , el tamaño mínimo del microsatélite sobre el cual puede ocurrir una mutación el cual se denominó “*delfloor*” y se fijó con un valor de 3, la estructura del ancestro común más reciente (*MRC*A), -incluyendo microsatélites sencillos, compuestos y ligados-, y el número de repeticiones (s) para obtener un promedio y varianza de los parámetros calculados, establecido por lo general como $s = 2000$. Es posible generar una muestra individual de haplotipos fijando $s = 1$.

3.1.5 *Parámetros calculados.* Realizando simulaciones con $s = 2000$ duplicados para cada grupo de parámetros iniciales, se obtuvo el promedio y la respectiva varianza de: el número de variantes y haplotipos, la heterocigosis¹ (Nei, 1978) y distintas apreciaciones de la homoplasia molecular. La homoplasia calculada a partir de la heterocigosis se estimó de acuerdo a Estoup *et al.* (2002). La homoplasia basada en el coalescente se calculó reconociendo como no homoplásticas a los haplotipos determinados por el conjunto de mutaciones más ancestrales; aquellas variantes que por tamaño convergieron con los haplotipos ancestrales se contemplaron como homoplásticas y tras su reconocimiento se calculó su proporción respecto al tamaño de la muestra.

Siguiendo el criterio anterior se determinó el orden de los haplotipos de acuerdo a su origen, del más ancestral al más derivado. Para cada grupo de parámetros se simularon 2000 duplicados y se registró el número de veces en que cada haplotipo, de acuerdo a su origen, fue el más frecuente de la muestra. Así, se obtuvo la proporción en que el haplotipo más ancestral fue también el más frecuente considerando el número de coincidencias respecto al número de duplicados.

El parámetro θ se empleó bajo dos aproximaciones: *a*) la θ esperada, correspondiente a $2N_e\mu$ y calculada a partir de los parámetros iniciales de cada simulación, y *b*) la θ calculada, la cual se computó de acuerdo a Kimura y Crown (θ_{IAM} ; 1964), siguiendo un modelo de alelos infinitos, y a Ohta y Kimura (θ_{OyK} ; 1973) y a Xu y Fu (θ_{XyF} ; 2004) para un modelo por pasos (SMM).

¹ $He = \frac{n}{n-1} (1 - \sum_i p_i^2)$, donde p es la frecuencia relativa de cada alelo y n es el tamaño de la muestra.

$$\theta_{IAM} = (1/F) - 1$$

$$\theta_{OyK} = [(1/F^2) - 1] / 2$$

$$F = \sum_i p_i^2,$$

donde p es la frecuencia relativa de cada alelo.

La θ_{XyF} por su parte, corresponde a un ajuste mediante a una regresión estadística de θ_{OyK} , y se calcula de la siguiente manera:

Cuando $\theta_{OyK} < 15$

$$\theta_{OyK} = (1.1313 + (3.4882 / n) + (28.2878 / n^2)) (\theta_{XyF}) + (0.3998)(\sqrt{\theta_{XyF}});$$

Cuando $\theta_{OyK} > 15$

$$\theta_{OyK} = (1.1675 + (3.3232 / n) + (63.698 / n^2)) (\theta_{XyF}) + (0.2569)(\sqrt{\theta_{XyF}})$$

donde n = tamaño de la muestra.

3.2 Redes de haplotipos

Las redes de haplotipos y/o variantes se construyeron con el programa TCS V. 1.21 (Clement *et al.*, 2000). Cada red se generó a partir de datos tomados de muestras simuladas individualmente, con diferentes combinaciones de parámetros y con diferentes configuraciones de muestra, particularmente con variantes multilocus y haplotipos. El haplotipo más ancestral se determinó por ser aquel con el conjunto de mutaciones más antiguas o bien aquel que no sufrió mutaciones a través de la genealogía. En el caso de las redes de variantes multilocus, la variante más ancestral se eligió: *a*) al no presentar homoplasia, igual que los haplotipos o *b*) al presentar homoplasia, como la variante que incluye dentro de los haplotipos que engloba al más antiguo de la muestra.

3.2.1 *Medidas aplicadas a las redes.* Se midieron ciertas medidas básicas correspondientes a la teoría de redes (Diestel, 2000):

- i. Número de nodos (también denominado *orden*).
- ii. Número de *hojas*: número nodos cuyo grado (numero de conexiones) es igual a 1.
- iii. Diámetro: distancia más grande entre un par de nodos, medida como el número de nodos que uno de ellos tiene que recorrer en el camino para llegar a la posición del otro.

IV. RESULTADOS

4.1 Subconjuntos de datos: electromorfos, variantes multilocus y haplotipos

Los microsatélites contenidos en las muestras producidas por las simulaciones se pueden agrupar en tres subconjuntos de acuerdo a diferentes aproximaciones estructurales:

- 1) Electromorfos o variantes por tamaño total: son aquellas que se distinguen únicamente por el tamaño del microsatélite obtenido después de una electroforesis.
- 2) Variantes multilocus: son las variantes cuya secuencia es revelada, de manera que se pueden distinguir casos en los que variantes idénticas por tamaño total corresponden en realidad a secuencias distintas.
- 3) Haplotipos por identidad o alelos: se reconocen sólo a través de las simulaciones ya que por medio de éstas se rastrea el lugar y el momento en que ocurren las mutaciones que conforman su estructura, dando así una identidad única a cada alelo, aunque pueda coincidir en tamaño total o secuencia multilocus con otros.

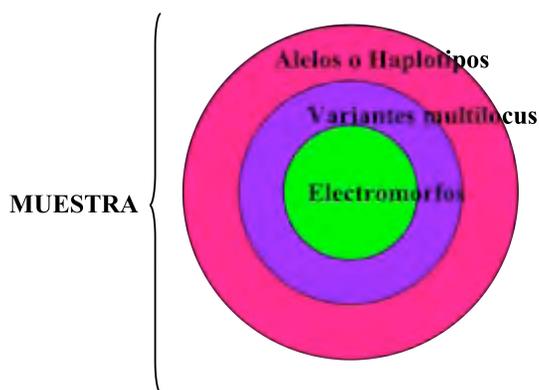


Figura 4.1 Diagrama de una muestra presentando la relación entre sus tres diferentes subconjuntos.

Así, estas tres configuraciones están contenidas en una muestra (Fig. 4.1) y varían entre sí numéricamente por la cantidad de elementos que se pueden reconocer en cada clase. Por ejemplo, si se consideran tres haplotipos de un microsatélite compuesto de dos loci:

- 1) AAAAGGGG
- 2) A*AAGGGGG
- 3) AAAGGGGG*

todos migrarán igualmente en un gel y por tanto, sólo se reconoce una variante por tamaño total. Al analizar su secuencia se puede descomponer la muestra en dos variantes multilocus, la primera representada por el alelo 1 y la segunda por los alelos 2 y 3. Ahora bien, si por medio de simulaciones, se encuentra que los haplotipos 2 y 3 son producto de mutaciones distintas (señaladas con asterisco) a pesar de compartir la misma secuencia, la muestra entonces revela 3 alelos distintos. De esta manera es evidente que la medida de variación genética está vinculada al subconjunto de la muestra con el que se está trabajando. Por ejemplo, en un caso muy sencillo como el anterior, donde el tamaño de la muestra es 3 considerando los tres haplotipos mostrados, la heterocigosis de los electromorfos es 0, la de las variantes multilocus 0.444 y la de los haplotipos 0.667.

Así, el valor de los parámetros calculados varía substancialmente dependiendo de la aproximación estructural que se esté considerando, razón por lo que en adelante se especificará con cuál subconjunto de la muestra se está trabajando.

4.2 Heterocigosis

La heterocigosis (Nei, 1987), es una medida que señala el nivel de diversidad genética en una muestra considerando el número y frecuencia de alelos polimórficos en un locus dado, de manera que se puede plantear como la probabilidad de que dos secuencias elegidas al azar correspondan a diferentes tipos alélicos.

El tamaño efectivo de la población y la tasa de mutación sin duda influyen en la cantidad de alelos presentes en una muestra. Mientras más elevados sean ambos parámetros, mayor será la heterocigosis (Figs. 4.2 y 4.3). El tamaño de la muestra por su

parte, no altera en gran medida los valores de heterocigosis calculados para cada tamaño efectivo y tasa de mutación.

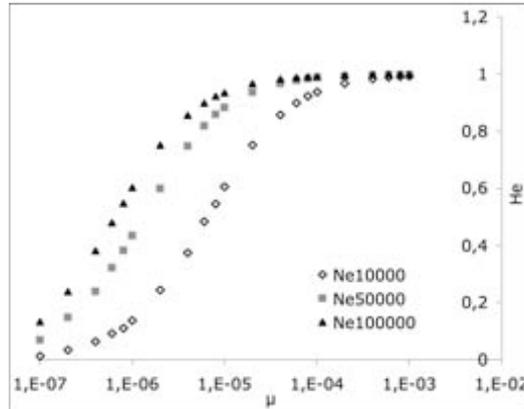


Figura 4.2 Medidas de heterocigosis de variantes multilocus para un microsatélite compuesto de 8 loci ligados (MRCA: T(11)-T(8)-A(8)-A(9)-T(15)-T(7)-A(9)-T(13)), con tamaños efectivos de las poblaciones de 10000, 50000 y 100000 y tamaño de muestra $k=100, 200, 500$ y 750 dentro de un intervalo de tasas de mutación de $\mu=10^{-7}$ a $\mu=10^{-3}$.

Como se mencionó anteriormente, la heterocigosis se calculó para los tres subconjuntos de la muestra: la heterocigosis de los electromorfos o H_E^T , que contempla las variantes por su tamaño total como un locus único, la heterocigosis multilocus o H_E^M que considera a las variantes multilocus y por último, la heterocigosis por identidad o H_E^I que se calcula contemplando los alelos verdaderos, conocidos únicamente a través de simulaciones y determinados por el conjunto de sus mutaciones.

Por su estructura, un microsatélite compuesto de dos o más loci presenta mayor posibilidad de variación, por tanto alcanza valores de heterocigosis (locus, multilocus y por identidad) mayores con respecto a aquellos sencillos. Así, mientras mayor es el tamaño total del microsatélite, el número de loci considerados o la cantidad de diferentes mutaciones que dan lugar a los alelos, mayor es la heterocigosis y ésta llega a su valor máximo ($H_E = 1$) a tasas de mutación más bajas en comparación con microsatélites menos complejos (Figs. 4.3 y 4.4).

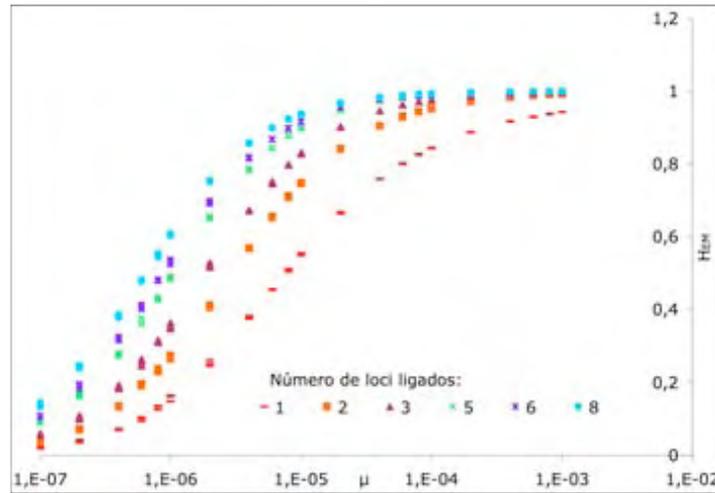


Figura 4.3 Heterocigosis multilocus para microsatélites de distinta composición, de 1 a 8 loci, $N_e=100000$, $k=100, 200, 500$ y 750 , dentro de un intervalo de tasas de mutación de $\mu=10^{-7}$ a $\mu=10^{-3}$.

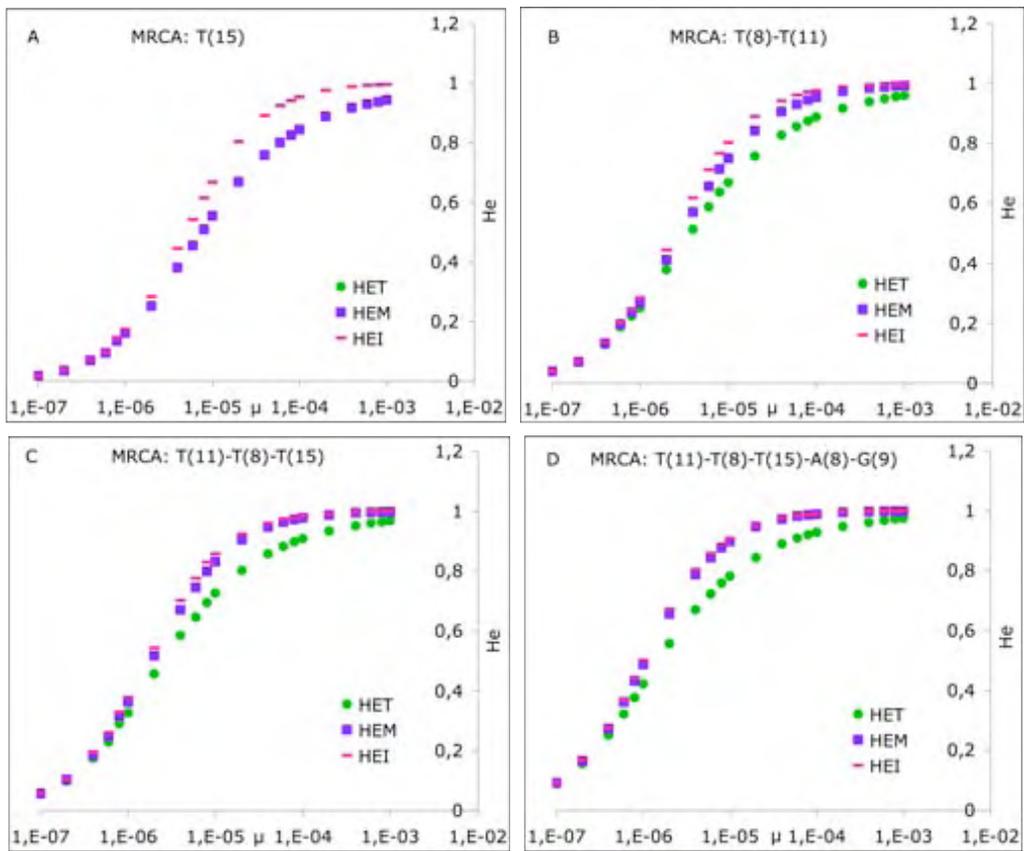


Figura 4.4 Comparación de las distintas medidas de heterocigosis respecto a la tasa de mutación: heterocigosis por tamaño total (H_E^T), heterocigosis multilocus (H_E^M) y heterocigosis por identidad (H_E^I) para microsatélites de diferente estructura, de 1 a 5 loci ligados (A-D). $N_e= 100000$; $k=250$; $s=2000$.

Por otra parte, respecto a la relación que guardan entre sí las tres medidas, la heterocigosis por tamaño total H_E^T , es menor que bajo una aproximación multilocus o por identidad. No obstante, la heterocigosis multilocus H_E^M arroja valores que se ajustan de manera muy cercana a la heterocigosis calculada por identidad (Fig. 4.4, A, B, C y D) salvo para el microsatélite simple, donde evidentemente la heterocigosis por tamaño total y multilocus son iguales ya que en ambos acercamientos se trata de un locus único (Fig. 4.4, A).

4.3 Número de variantes y haplotipos

El número de variantes o haplotipos es otro parámetro que refleja la diversidad genética.

Al igual que en la heterocigosis, una tasa de mutación elevada produce mayor cantidad de haplotipos o variantes con respecto a tasas de mutación más lentas. Igualmente, un mayor tamaño efectivo de la población genera mayor diversidad genética, la cual, por otra parte, se ve enriquecida mientras más loci compongan al microsatélite en cuestión (Fig. 4.5). La tasa relativa de mutación, θ ($\theta = 2Ne\mu$), es el parámetro que engloba estos tres elementos centrales en la determinación de la variación genética de los microsatélites (la tasa de mutación, el tamaño efectivo de la población y el número de loci); por ello, en adelante, muchos de los resultados se ejemplificarán en función de dicho parámetro, excepto cuando se quiera enfatizar el efecto que tienen por separado sus componentes.

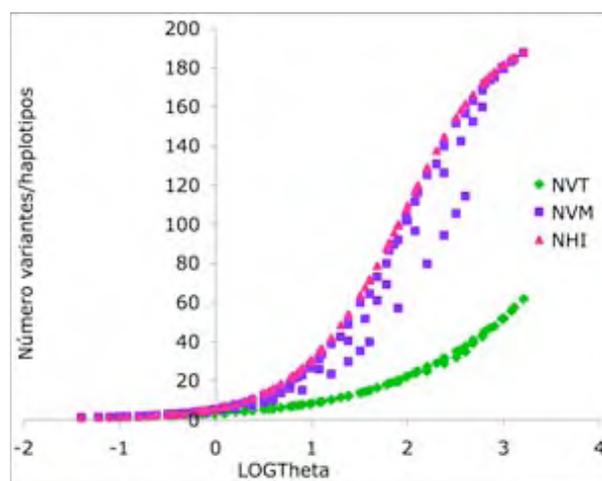


Figura 4.5 Número de variantes por tamaño total (N_V^T), variantes multilocus (N_V^M) y haplotipos (N_H^I), respecto a θ ($2N_e\mu$), para microsatélites de distinta composición, de 2 a 8 loci ligados.

El número de electromorfos es el más reducido. Por su parte, el número de variantes multilocus y de haplotipos es acentuadamente mayor. Así, los haplotipos presentan siempre la mayor diversidad, seguidos de las variantes multilocus y por debajo de ambas, las variantes por tamaño total o electromorfos (Fig. 4.5). No obstante, conforme más loci se suman a la estructura del microsatélite, se desarrolla una creciente semejanza entre el número de haplotipos y el número de variantes multilocus, de manera que mientras para un microsatélite compuesto de dos loci el número de variantes multilocus se mantiene como un intermedio entre el número de variantes por tamaño total y el número de haplotipos, para un microsatélite de más de cuatro loci la diferencia entre la cantidad de variantes multilocus y de haplotipos es casi indistinguible (Fig. 4.6).

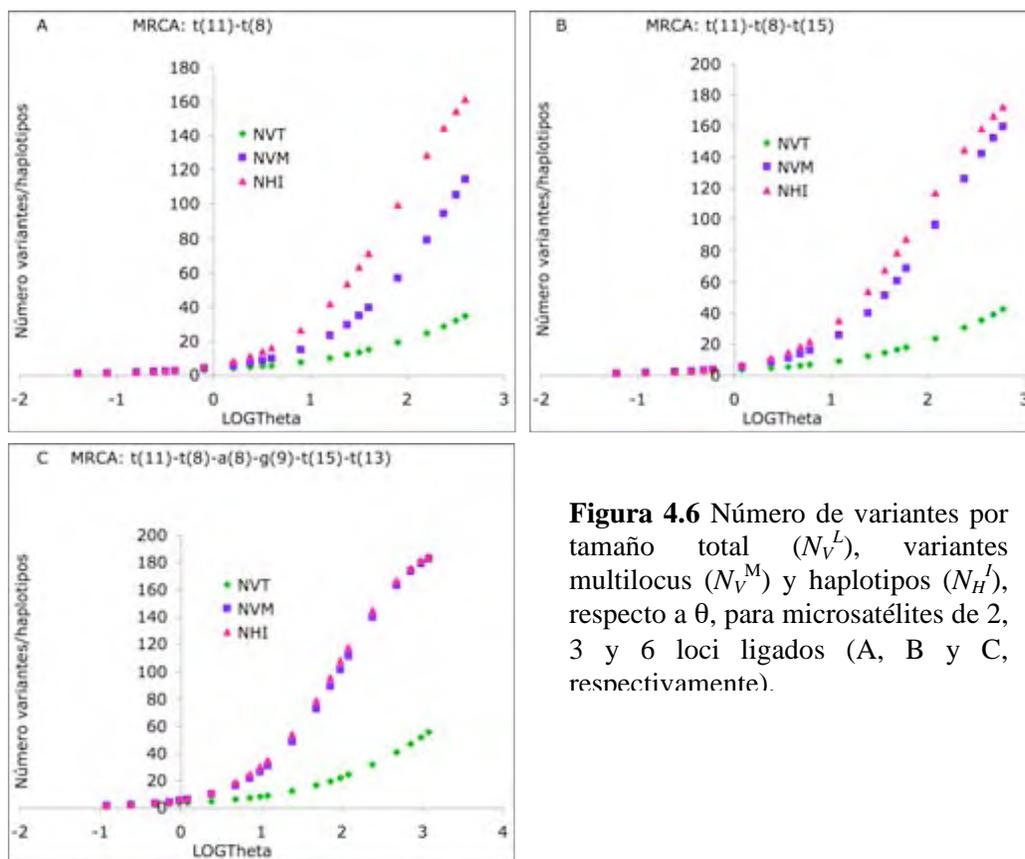


Figura 4.6 Número de variantes por tamaño total (N_V^L), variantes multilocus (N_V^M) y haplotipos (N_H^L), respecto a θ , para microsatélites de 2, 3 y 6 loci ligados (A, B y C, respectivamente).

Cabe señalar que el tamaño de la muestra es un parámetro que influye en el conteo de variantes o haplotipos, de manera que cuando el tamaño de la muestra es demasiado pequeño respecto al tamaño efectivo de la población se subvalora en gran medida el

número de variantes o haplotipos, particularmente para valores de θ^l superiores a 10 (Fig. 4.7, A).

La heterocigosis por su parte, no es un parámetro que sea tan sensible al tamaño de la muestra como se vio anteriormente (Figs. 4.2 y 4.3). Para valores de heterocigosis menores a 0.8 el número de variantes y/o haplotipos es muy semejante para tamaños de muestra que oscilan entre $k= 25$ y $k= 750$ por lo que no es extraño que la heterocigosis mantenga valores muy constantes a pesar de dicha variación. No obstante, a partir de $H_E= 0.8$ y hasta el máximo $H_E= 1$, el número de variantes y/o haplotipos varía sustancialmente en relación con el tamaño de la muestra (Fig. 4.7, B).

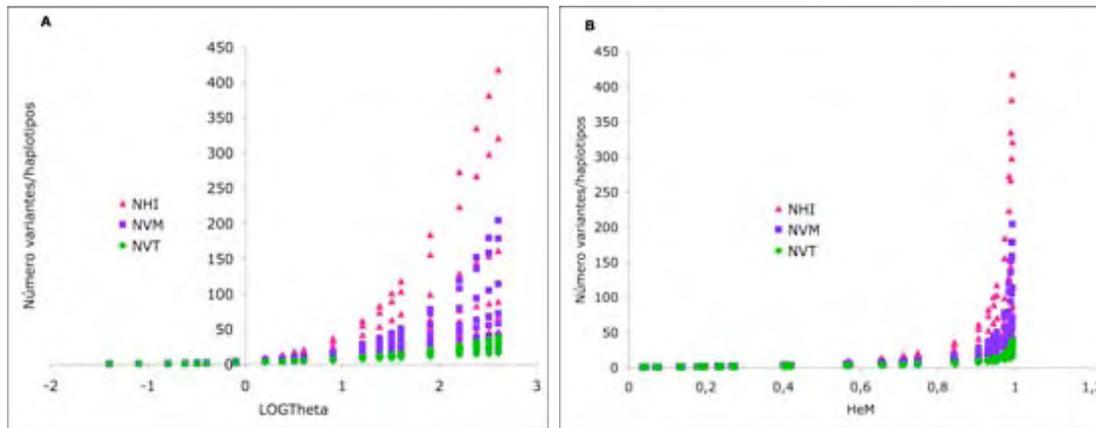


Figura 4.7 Número de variantes y haplotipos respecto a θ (A) y respecto a la heterocigosis multilocus (B) para un microsatélite con 2 loci ligados (MRCA: A(8)-G(9)), $k= 25, 50, 75, 100, 200, 500$ y 750 (en la figura, los números más altos de haplotipos o variantes para cada valor de θ y H_E^M corresponden al tamaño de muestra k más elevado y así sucesivamente), $N_e=100000$, $s=2000$.

Ya que la heterocigosis es una medida acotada a un valor máximo, 1, un número muy distinto de variantes/haplotipos pueden producir altos valores de heterocigosis, no obstante es importante considerar que aunque cierto número de variantes/haplotipos determinado por muy distintos tamaños de muestra pueden generar un valor de heterocigosis muy elevado, puede subvalorarse el número de variantes/haplotipos que representan a la población. Por ejemplo, en la figura 4.7B se muestra el número de variantes/haplotipos respecto a su

¹ La θ calculada tanto para las variantes multilocus como para los haplotipos por identidad, toma en consideración la variación de cada locus ligado por lo que se denomina θ multilocus. En adelante cada mención de este parámetro se refiere a su versión multilocus, tomando en consideración que las variantes por tamaño total (electromorfos) corresponden a un único locus.

heterocigosis multilocus H_E^M para distintos tamaños de muestra. Cuando la $H_E^M=1$ el número de haplotipos (N_H^J) oscila ampliamente de 25 hasta casi 420 dependiendo de el tamaño de la muestra, igualmente el número de variantes multilocus (N_V^M) que va de 22 a 205 y el de variantes por tamaño total (N_V^T), aunque de manera menos radical, de 16 a 40.

4.4 Homoplasia

La proporción de alelos homoplásticos, idénticos por estado y no por descendencia, se obtuvo bajo dos acercamientos distintos. El primero se basa en la medida de heterocigosis para cada subconjunto de la muestra y la relación que existe entre estos. El segundo es sólo posible mediante simulaciones y se fundamenta en una medición directa de alelos homoplásticos a partir del coalescente.

4.4.1 Homoplasia calculada a partir de la heterocigosis. La medida del índice de homoplasia (Estoup *et al.*, 2002) está basada en la probabilidad de que dos alelos sean idénticos por estado bajo un modelo de mutación paso a paso (*stepwise mutation model* o SMM) respecto a un modelo de alelos infinitos (*Infinite Allele Model* o IAM); ambas probabilidades son iguales a la homocigosis calculada con base en ambos modelos, es decir, $(1 + 2M)^{-0.5}$ para SMM y $(1 + M)^{-1}$ para IAM ($M = 2N_e\mu$) donde dos copias de un gen son iguales únicamente si provienen del mismo ancestro sin haber mutado. La proporción entre la homocigosis IAM sobre la homocigosis SMM expone la probabilidad de que dos genes sea idénticos por estado y por descendencia, así pues:

$$S = 1 - [(1 + 2M)^{0.5}/1 + M]$$

corresponde a la proporción de alelos homoplásticos o idénticos por estado mas no por descendencia de acuerdo a Estoup y colaboradores (2002).

De esta manera y a partir de las simulaciones se obtienen tres valores de heterocigosis: la heterocigosis por identidad (H_E^J), la heterocigosis por tamaño total (H_E^T) y la heterocigosis multilocus (H_E^M). La homocigosis bajo un modelo de alelos infinitos

corresponde a $1-H_E^I$, donde aquellos alelos idénticos por estado son necesariamente idénticos por descendencia. Por otra parte, la homocigosis bajo un modelo paso a paso corresponde a la homocigosis por tamaño, ya sea locus o multilocus ($1-H_E^T$ o $1-H_E^M$, respectivamente), donde las variantes pueden coincidir en tamaño a pesar de provenir de ancestros inmediatos distintos.

Así, el índice de homoplasia de Estoup define la proporción de homocigosis por identidad respecto a la homocigosis por tamaño, de manera que calcula la homoplasia de tamaño total o locus (*size homoplasia* o *SH*) y la homoplasia multilocus o estructuralmente accesible (*structurally accesible size homoplasia* o *SASH*):

$$SH = 1 - [(1 - H_E^I)/(1 - H_E^T)]$$

$$SASH = 1 - [(1 - H_E^I)/(1 - H_E^M)]$$

Ambas medidas de homoplasia, *SH* y *SASH*, son substancialmente diferentes. Bajo ambas apreciaciones, conforme aumenta la tasa de mutación incrementa la proporción de alelos idénticos por estado y no así por descendencia, sin embargo, la tendencia de incremento es muy distinta en los dos casos (Fig. 4.8, A y B). Todos los electromorfos tienden a 100% de homoplasia de tamaño (*SH*) hacia la tasa de mutación más elevada ($\mu=10^{-3}$) (Fig. 4.8, A). No obstante, considerando una aproximación por secuencia o multilocus, la composición de cada microsatélite interviene de manera fundamental en la proporción de alelos homoplásticos. Un microsatélite sencillo, compuesto de una secuencia de nucleótidos idénticos, presenta valores de homoplasia estructuralmente accesible (*SASH*) muy elevados con respecto a aquellos más complejos; mientras más loci se sumen a la estructura de un microsatélite, menor será el nivel de homoplasia que se pueda presentar entre sus alelos. No obstante, a tasas de mutación por debajo de $\mu=2 \times 10^{-6}$, la proporción de *SASH* que presentan los microsatélites de distinta estructura es muy similar y se mantiene por debajo de 10% (Fig. 4.8, B).

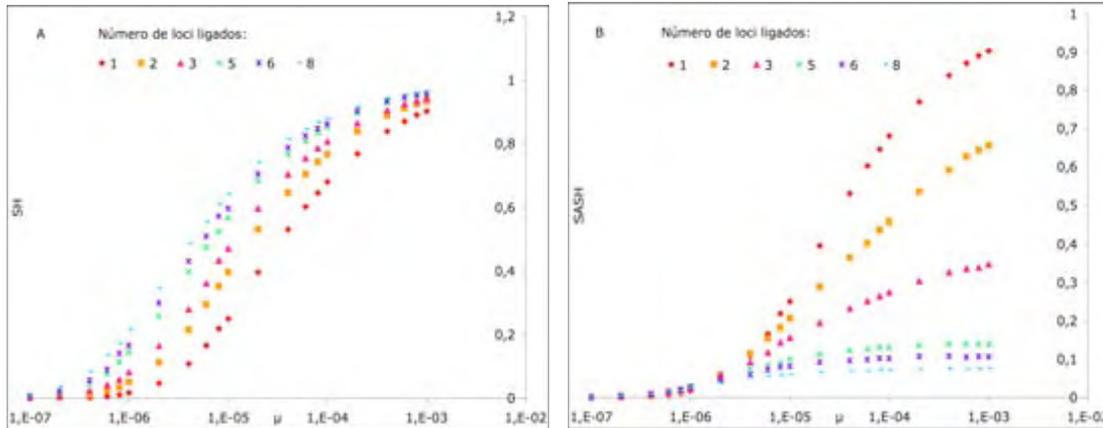


Figura 4.8 Homoplasiya por tamaño total, *SH* (A) y homoplasiya multilocus o estructuralmente accesible, *SASH* (B), para microsatélites de distinta composición, de 1 a 8 loci ligados, dentro de un intervalo de tasas de mutación de $\mu=10^{-7}$ a $\mu=10^{-3}$, con $N_e=100000$ y $k=200$.

Una tercera medida de homoplasiya basada en la heterocigosis es posible. Corresponde a la homoplasiya molecularmente accesible o *MASH*, propuesta también por Estoup *et al.* Dicha medida adquiere su relevancia porque su cálculo es posible con datos obtenidos en el laboratorio, ya que se basa en la comparación de la homocigosis multilocus respecto a la homocigosis por tamaño total.

La homoplasiya molecularmente accesible (*MASH*) se abordará con mayor profundidad en el apartado 4.4.3.

4.4.2 Homoplasiya basada en el coalescente. La homoplasiya basada en el coalescente (*PC*) calcula los alelos homoplásticos a partir de una genealogía reconstruida mediante simulaciones, registrando las mutaciones que dan origen a la muestra de variantes y el momento en que surgen. Con este fundamento, la homoplasiya basada en el coalescente arroja un valor preciso de la homoplasiya encontrada en un linaje simulado de haplotipos (Fig. 4.9).

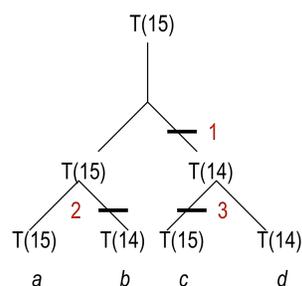


Figura 4.9 Ejemplo de una genealogía con las mutaciones correspondientes que dan origen a una muestra de haplotipos en el presente. El seguimiento en el tiempo permite ubicar los tipos ancestrales necesarios para calcular la homoplasiya coalescente.

De acuerdo a la genealogía representada en la figura 4.9, tres mutaciones dan origen a la muestra de cuatro haplotipos distintos por identidad (a , b , c y d). Por su tamaño, estos se agrupan en dos subconjuntos: T(15): a y c , y T(14): b y d , de manera que en cada subconjunto se reúne un haplotipo homoplástico respecto al otro, siendo el no homoplástico el más ancestral o bien para nuestro ejemplo, siendo c homoplástico respecto a a , y b homoplástico respecto a d .

Siguiendo este principio se distinguen los haplotipos no homoplásticos en una genealogía y es posible calcular la proporción de aquellos que sí lo son. Por ejemplo, para el caso expuesto en la figura 4.9 la proporción de homoplasia es $2/4$ o bien el 50%.

Así, la homoplasia basada en el coalescente se puede definir como:

$$PC = \text{No. de haplotipos homoplásticos} / k,$$

donde k = tamaño de la muestra.

Los valores de homoplasia calculada a partir de la heterocigosis bajo sus distintas aproximaciones (SH , $SASH$ y $MASH$, esta última se abordará más adelante) se pueden estudiar en relación a la homoplasia basada en el coalescente (PC_{SH} , PC_{SASH} y PC_{MASH}) la cual, como se explicó, mide la proporción de homoplasia de manera directa en cada genealogía y provee un punto de comparación. Para valores de θ entre 0.5 y 10 ambas medidas son bastante similares, salvo por una ligera subvaloración de SH , $SASH$ y $MASH$. A partir de $\theta > 10$ la homoplasia basada en la heterocigosis y la coalescente se hacen dispares, de manera que mientras la primera sigue una tendencia ascendente, la segunda llega un máximo, que varía de acuerdo al tamaño de muestra, y desciende (Fig. 4.10).

Dado a que la formulación del índice de homoplasia SH , $SASH$ y $MASH$ se fundamenta en el cálculo de la heterocigosis de dos subconjuntos de datos y por lo mencionado arriba los parámetros decisivos son la tasa de mutación y el tamaño efectivo de la población, el tamaño de la muestra (k) no altera el valor de homoplasia obtenido. No obstante, la proporción de alelos homoplásticos, contados a partir del coalescente, es distinta para cada tamaño de muestra. Conforme el tamaño de la muestra aumenta, menor

es la diferencia entre la homoplasia basada en la heterocigosis y la homoplasia basada en el coalescente.

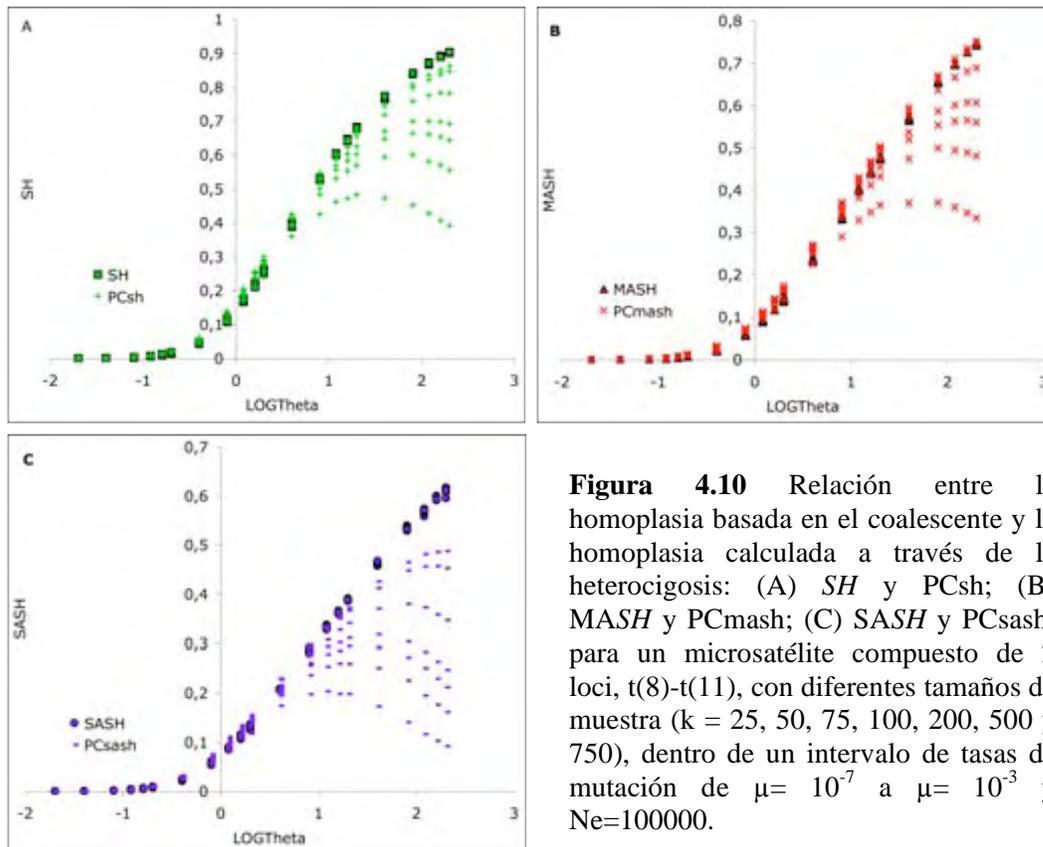


Figura 4.10 Relación entre la homoplasia basada en el coalescente y la homoplasia calculada a través de la heterocigosis: (A) *SH* y *PCsh*; (B) *MASH* y *PCmash*; (C) *SASH* y *PCsash*; para un microsatélite compuesto de 2 loci, t(8)-t(11), con diferentes tamaños de muestra ($k = 25, 50, 75, 100, 200, 500$ y 750), dentro de un intervalo de tasas de mutación de $\mu = 10^{-7}$ a $\mu = 10^{-3}$ y $N_e = 100000$.

4.4.3 Homoplasia molecularmente accesible. El cálculo de las medidas de homoplasia mencionadas hasta ahora es posible únicamente mediante simulaciones en la computadora. Sólo a través de éstas es posible obtener información que no se puede generar experimentalmente como la heterocigosis por identidad, necesaria para calcular la homoplasia por tamaño (*SH* y *SASH*), y el rastreo de las mutaciones a lo largo de un linaje de haplotipos para calcular la homoplasia coalescente.

Una tercera medida de homoplasia, la homoplasia molecularmente accesible (Estoup *et al.*, 2002), es fundamental ya que se calcula con datos que se pueden obtener experimentalmente en el laboratorio. Esta medida corresponde a una fracción de la homoplasia de tamaño y se detecta cuando, mediante técnicas moleculares como la secuenciación, dos variantes idénticas por tamaño revelan secuencias distintas.

La homoplasia molecularmente accesible (*molecularly accessible size homoplasia* o *MASH*) es pues una razón de la proporción de variantes homoplásticas por secuencia o estructuralmente accesibles respecto a las homoplásticas por tamaño (Estoup *et al.*, 2002). En otras palabras, permite calcular la fracción de homoplasia por tamaño que es posible discriminar mediante el conocimiento de diferencias en secuencia de los electromorfos.

Con base en lo anterior y retomando el fundamento de la homoplasia calculada a partir de la heterocigosis, se puede definir a la homoplasia molecularmente accesible como la proporción de homocigosis por tamaño multilocus respecto a la homocigosis por tamaño total, de manera que:

$$MASH = 1 - [(1 - H_E^M)/(1 - H_E^T)]$$

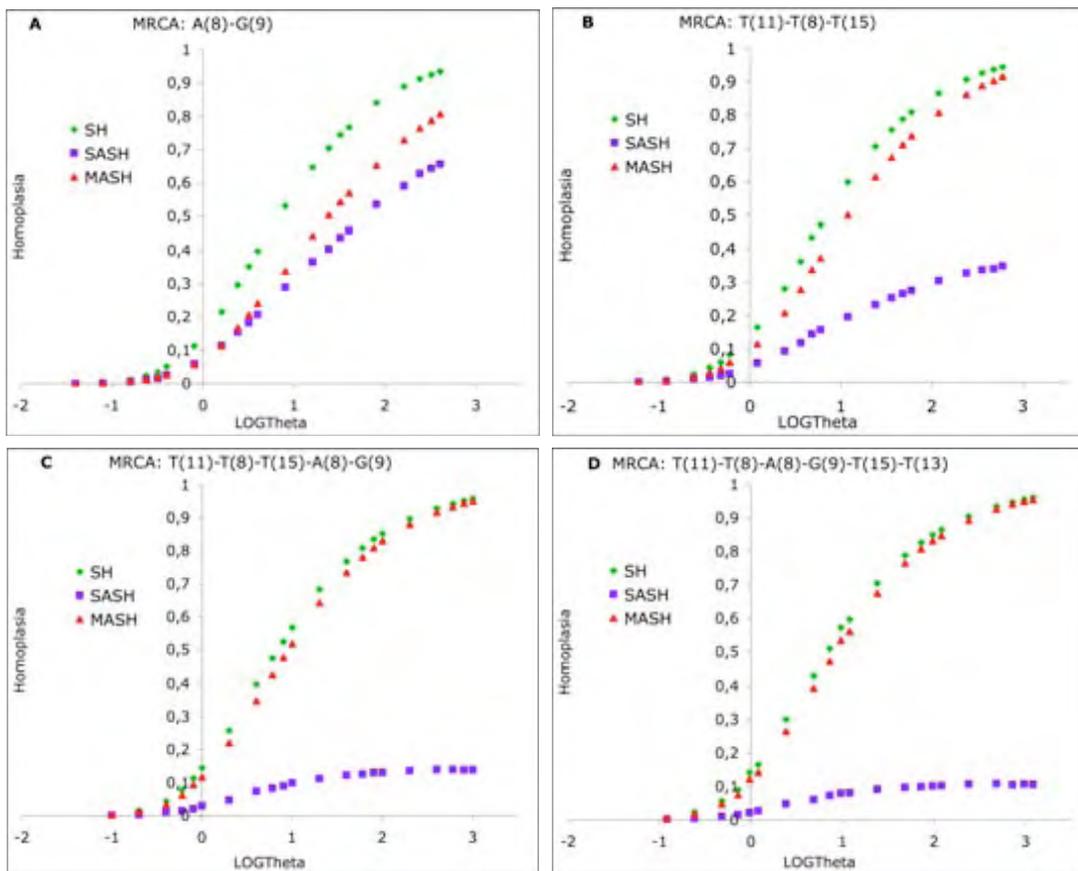


Figura 4.11 Relación de los tres distintos acercamientos a la homoplasia calculada por medio de la heterocigosis respecto a θ : homoplasia por tamaño total (*SH*), homoplasia multilocus (*SASH*) y homoplasia molecularmente accesible (*MASH*), para 2, 3, 5 y 6 microsatélites ligados (A, B, C y D, respectivamente). $N_e=100000$ y $k=200$.

La composición del microsatélite, respecto al número de loci, influye en la medida que se obtiene de la homoplasia molecularmente accesible y por tanto la relación que ésta tiene con la homoplasia por tamaño total.

En la figura 4.11 se expone la relación que poseen entre sí las tres medidas de homoplasia basadas en la heterocigosis: *SH*, *SASH* y *MASH*. Los valores de la homoplasia molecularmente accesible se aproximan a los de la homoplasia por tamaño total a medida que más loci componen al microsatélite. De modo que:

$$SH \approx MASH$$

Es decir,

$$1 - [(1 - H_E^I)/(1 - H_E^T)] \approx 1 - [(1 - H_E^M)/(1 - H_E^T)]$$

Para que dicha condición se cumpla, la heterocigosis por identidad y la heterocigosis multilocus deben adquirir valores similares:

$$H_E^I \approx H_E^M$$

De manera que cuando tres o más loci se reconocen en la estructura de un microsatélite, la heterocigosis por identidad y la heterocigosis multilocus son aproximadamente iguales (Fig. 4.4). A partir de esto se explica la reducción de la homoplasia estructuralmente accesible respecto al número de loci que estructuran a un microsatélite.

Si,

$$H_E^I \approx H_E^M$$

entonces,

$$1 - H_E^I \approx 1 - H_E^M$$

De acuerdo a las simulaciones, cuando se tienen tres o más loci, la relación

$$1-H_E^I/1-H_E^M > 0.86$$

de manera que

$$1-[(1-H_E^I)/(1-H_E^M)] < 0.14$$

es decir,

$$SASH < 14\%$$

No obstante cabe recalcar que mientras H_E^I y H_E^M no sean iguales, es decir, mientras exista una diferencia entre ambas medidas, por más pequeña que sea, la homoplasia estructuralmente accesible nunca será igual a 0.

4.5 Homoplasia y heterocigosis

La relación que existe entre la heterocigosis y la homoplasia permite referir la proporción de alelos idénticos por estado mas no por descendencia a una medida de variabilidad genética.

Mientras mayor es la heterocigosis, mayor es la proporción de todos los tipos de homoplasia. El valor de esta última se ve afectado por la complejidad del microsatélite, de manera que aquellos compuestos de dos o más secuencias distinguibles entre sí reducen considerablemente la probabilidad de generar variantes homoplásticas (Fig. 4.12).

Si, por ejemplo, se considera un microsatélite sencillo, a partir de valores de heterocigosis mayores a 0.7, más de 50% de las variantes serán homoplásticas. Por otra parte, para un microsatélite constituido por dos loci, esta proporción se reduce; a una heterocigosis multilocus de 0.7, como en el caso anterior, la proporción de homoplasia no rebasa el 20%. Así, mientras más loci son considerados, menor es el nivel de homoplasia (Fig. 4.12).

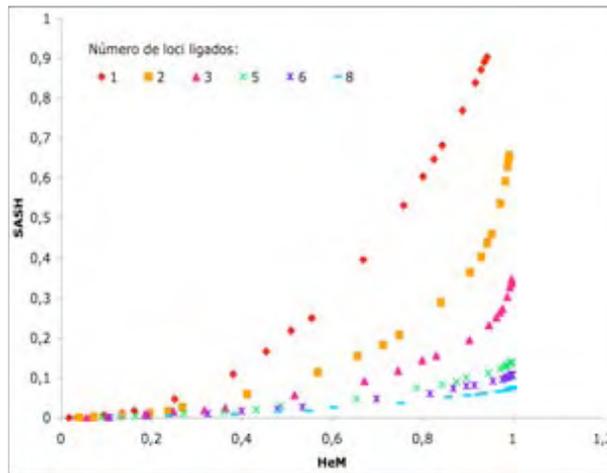


Figura 4.12 Homoplasia estructuralmente accesible (*SASH*) calculada a partir de la heterocigosis para microsatélites de distinta composición, en relación con la heterocigosis multilocus. $N_e=100000$ y $k=200$.

4.6 Homoplasia y tamaño efectivo de la población

Otro parámetro que afecta la medida tanto de heterocigosis como de homoplasia es el tamaño efectivo de la población (N_e).

Retomando el planteamiento del índice de homoplasia de Estoup *et al.* (2002), donde la proporción entre la homocigosis calculada bajo un modelo de mutación de alelos infinitos sobre la homocigosis calculada bajo un modelo de mutación paso a paso expone la probabilidad de que dos alelos sean idénticos por estado, mas no por descendencia, y dicha proporción corresponde a $1 - [(1 + 2M)^{0.5}/(1 + M)]$, se puede apreciar que el índice de homoplasia, calculado de esta manera, depende del parámetro $M = 2N_e\mu$, donde N_e es el tamaño efectivo de la población y μ la tasa de mutación. A partir de esto es evidente que N_e y μ tienen un papel central en las medidas de heterocigosis y homoplasia, de manera que a mayor tamaño de la población y a tasas de mutación más elevadas corresponden valores más altos para ambas medidas.

Ahora bien, de acuerdo a la teoría de coalescencia, una muestra de alelos de una población grande tomará más tiempo en coalescer a su ancestro común más reciente, que la misma muestra tomada de una población comparativamente más pequeña. Mientras más

grande es la población, mayor es el tiempo a la coalescencia y por tanto, mayor es el número de generaciones que acumulan mutaciones a una tasa de mutación dada. De esta manera se espera encontrar una mayor variedad de alelos en una población grande, es decir mayor variedad genética traducida en mayor heterocigosis, y éstos tendrán mayor probabilidad de converger por estado.

Ejemplificando lo anterior, un microsatélite compuesto por dos secuencias, expone una proporción de homoplasia por encima de 45% cuando se considera una tasa de mutación elevada, $\mu = 10^{-3}$, alcanzando su máximo en el mayor tamaño de la población, es decir, 100,000 individuos, donde casi 70% de las variantes son homoplásticas. Así mismo, para las tasas de mutación $\mu = 10^{-4}$ y 10^{-5} , el tamaño efectivo de la población tiene un efecto trascendente en el nivel de homoplasia, es decir, si se evalúa *SASH* a una tasa de mutación de $\mu = 10^{-4}$, para un tamaño efectivo de 10,000 apenas rebasa 0.2, sin embargo para $N_e = 100,000$ llega a 0.46 lo cual corresponde ya a una proporción relevante. A $\mu = 10^{-5}$ la proporción máxima de homoplasia no alcanza el 30%, mas es un valor relativamente alto si se compara con el 2.5% correspondiente a un tamaño efectivo de 10,000. Para el resto de las tasas de mutación ocurre el mismo comportamiento, no obstante son tan lentas que, aun para tamaños efectivos grandes, la homoplasia no alcanza valores significativos manteniéndose por debajo de 2.2% de ocurrencia (Fig. 4.13, A).

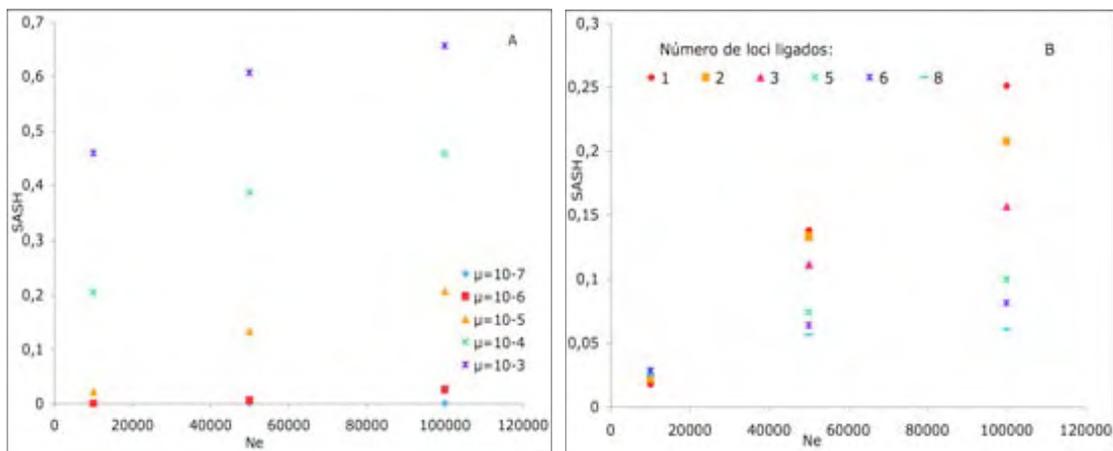


Figura 4.13 Homoplasia estructuralmente accesible (*SASH*) calculada a partir de la heterocigosis respecto al tamaño efectivo poblacional para un microsatélite compuesto de dos loci a diferentes tasas de mutación (A) y para microsatélites de diferente composición, de 1 a 8 loci ligados, a una tasa de mutación de $\mu = 10^{-6}$ (B). $k=200$.

Dicho comportamiento se mantiene con microsatélites de distintas complejidades, es decir, la proporción de alelos homoplásticos es mayor cuando se trata de tamaños efectivos poblacionales grandes. Sin embargo, como se planteó arriba, mientras más loci componen al microsatélite menor es la homoplasia estructuralmente accesible (*SASH*) que presenta.

Esto es evidente al comparar el índice de homoplasia para microsatélites de distinta composición a una misma tasa de mutación ($\mu=10^{-5}$ en este caso) y distintos tamaños poblacionales. Para un tamaño efectivo pequeño (10,000 individuos), todos los microsatélites muestran una probabilidad similar de presentar alelos homoplásticos, alrededor del 2.2%. No obstante, para tamaños efectivos mayores, de 50,000 o 100,000 individuos, por ejemplo, es claro que el microsatélite más simple presenta el mayor grado de *SASH*, y que ésta se reduce en función del número de secuencias distinguibles que estructuran al resto de los microsatélites (Fig. 4.13, B).

4.7 Homoplasia y θ

θ es un parámetro fundamental ya que brinda información sobre el tamaño efectivo de las poblaciones y la tasa de mutación, los cuales, como se menciona arriba, son responsables del grado de variabilidad genética.

El parámetro θ que se empleó en este apartado corresponde a la θ esperada, la cual corresponde a la definición [$2N_e\mu$ x (número de loci del microsatélite)], donde N_e = tamaño efectivo de las poblaciones y μ = tasa de mutación. Los parámetros empleados para calcular θ fueron los mismo que se utilizaron para realizar las simulaciones y se relacionaron a los valores que éstas últimas arrojaron.

La relación que guardan las distintas homoplasias basadas en la heterocigosis en función de θ , expone patrones claros del comportamiento de cada una de éstas y ofrece un punto de referencia e inferencia respecto al grado de homoplasia que se podría esperar considerando una θ dada.

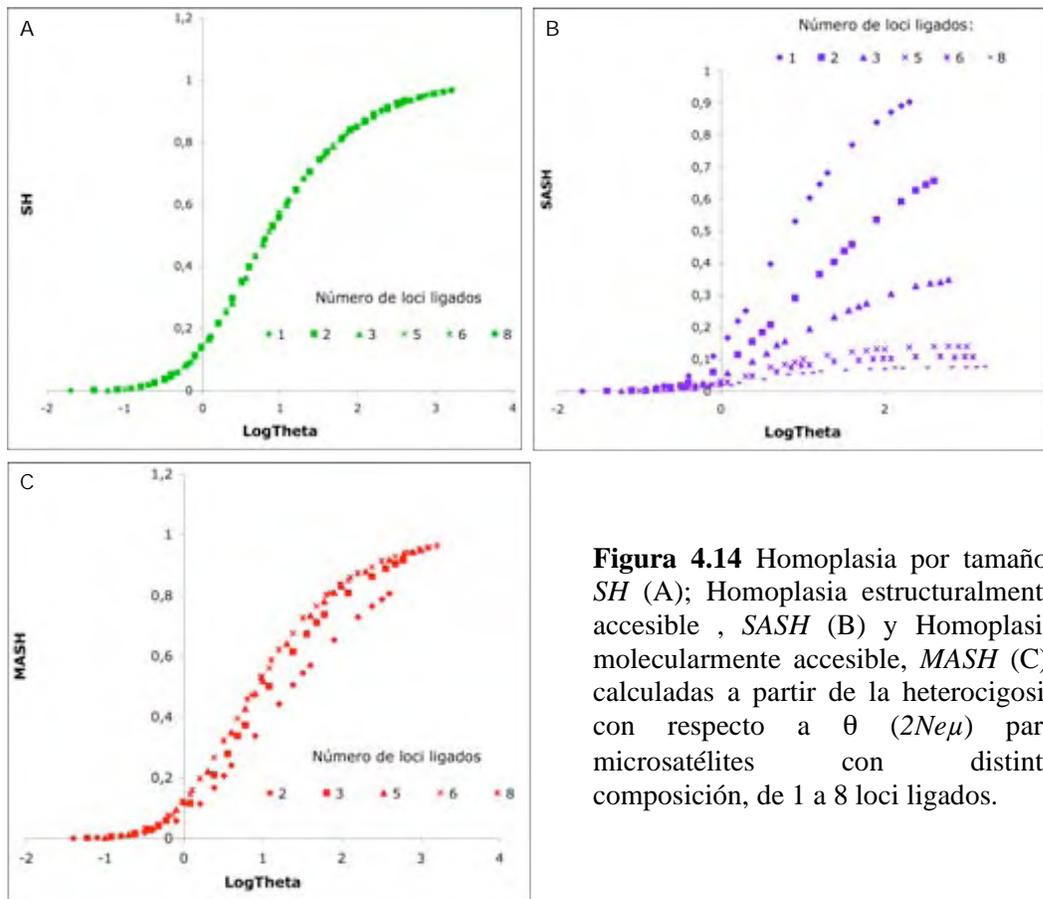


Figura 4.14 Homoplasia por tamaño, *SH* (A); Homoplasia estructuralmente accesible, *SASH* (B) y Homoplasia molecularmente accesible, *MASH* (C), calculadas a partir de la heterocigosis con respecto a θ ($2Ne\mu$) para microsatélites con distinta composición, de 1 a 8 loci ligados.

La homoplasia, en sus distintas apreciaciones, asciende conforme θ aumenta, sin embargo, cada una de ellas muestra un patrón diferente relacionado con la estructura del microsatélite. Por un lado, la homoplasia por tamaño, *SH*, no se ve afectada por la simplicidad o complejidad de los microsatélites sino que sigue una sola curva a lo largo de todos los valores de θ (Fig. 4.14, A). Para la homoplasia estructuralmente accesible, *SASH*, el número de loci contemplado se reveló como un factor sumamente trascendente, ya que a mayor número de loci, *SASH* se reduce considerablemente (Fig. 4.14, B). La homoplasia molecularmente accesible, *MASH*, muestra un comportamiento similar al de *SH*, no obstante sí evidencia un cierto efecto relacionado a la composición del microsatélite donde se aprecia que por debajo de 3 loci la accesibilidad molecular es ligeramente menor que para otros microsatélites más complejos, sin embargo, de 5 loci en adelante la accesibilidad molecular es muy similar (Fig. 4.14, C).

4.8 Haplotipos ancestrales y genealogías

Como se ha mencionado, cada muestra simulada de haplotipos incluye el conjunto de mutaciones que caracterizan a cada alelo y el orden en que dichas mutaciones ocurrieron, por tanto es posible ubicar al haplotipo más ancestral en cada caso y así mismo reconocer a aquellos más derivados.

Se ha propuesto que el haplotipo más ancestral de una muestra es aquel que presenta una frecuencia elevada y un alto grado de conectividad en una red de haplotipos, idealmente ocupando una posición central dentro de ésta (Posada y Crandall, 2001).

Las redes construidas a partir de simulaciones muestran que sólo ocasionalmente el haplotipo más ancestral coincide con el más abundante. La tasa de mutación y el tamaño efectivo de las poblaciones influyen esta proporción en gran medida. La probabilidad de que el haplotipo más ancestral sea el más abundante en una muestra, incrementa mientras menor sea el tamaño efectivo de las poblaciones y la tasa de mutación sea más lenta, condiciones que a su vez implican una heterocigosis y un nivel de homoplasia muy reducidas, por el contrario conforme la tasa de mutación y el tamaño efectivo se elevan, esta probabilidad decrece. Así, de acuerdo a las proporciones obtenidas de las simulaciones para un microsatélite con 6 loci ligados, a un tamaño efectivo de 50,000 individuos en adelante a una tasa de mutación de $\mu = 10^{-5}$, la probabilidad de que el haplotipo más ancestral sea el más abundante se encuentra por debajo de 22%. Ahora bien, esta probabilidad aumenta por arriba de 40%, aun para un tamaño efectivo alto de 300,000 cuando la tasa de mutación es de $\mu = 10^{-6}$. Para tasas de mutación mayores, como $\mu = 10^{-4}$, el haplotipo más ancestral no coincide con el más abundante ni siquiera con una proporción muy reducida (Fig. 4.15); son haplotipos más derivados los que predominan en frecuencia, sin embargo, de acuerdo a su origen, si bien no corresponden al ancestro, son alelos separados de él por pocas mutaciones, cuya aparición temporal no dista mucho de las que componen al más ancestral.

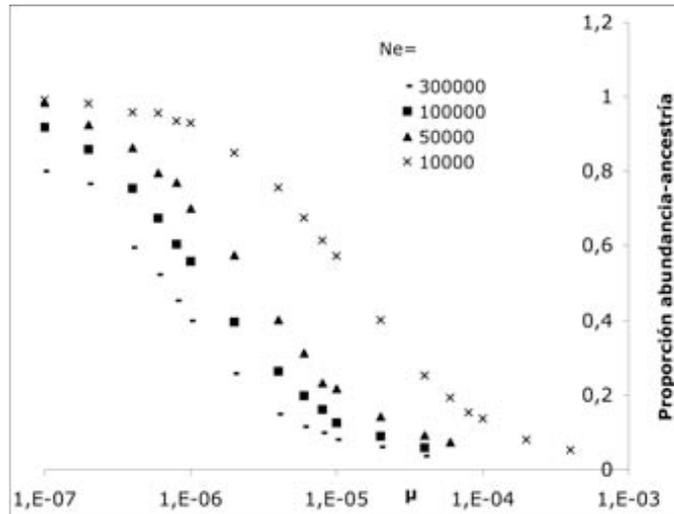


Figura 4.15 Probabilidad de que el haplotipo más ancestral es también el más abundante para distintos tamaños poblacionales dentro de un intervalo de tasas de mutación de $\mu=10^{-7}$ a $\mu=10^{-3}$, para un microsatélite de 6 loci ligados.

Más aún, la probabilidad con la que el haplotipo más ancestral es también el más frecuente varía con el número de loci que componen al microsatélite; dicha probabilidad decrece a medida que se presentan más loci ligados (Fig. 4.16).

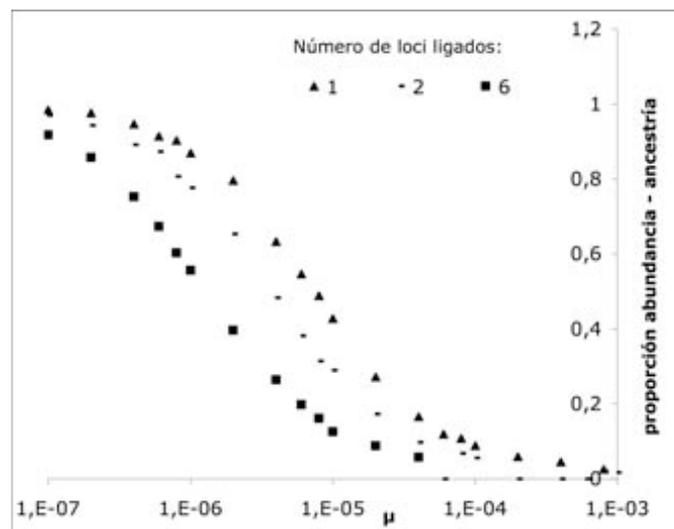


Figura 4.16 Probabilidad de que el haplotipo más ancestral es también el más abundante para dos microsatélites de distinta composición (1 locus, 2 loci y 6 loci ligados) dentro de un intervalo de tasas de mutación de $\mu= 10^{-7}$ a $\mu= 10^{-3}$, $N_e= 100000$.

El parámetro θ ofrece otro punto de análisis a este respecto. θ refleja la relación existente entre la tasa de mutación, el tamaño efectivo poblacional y el número de loci ligados que se evalúan en cada caso (Fig. 4.17). La proporción en que el haplotipo más ancestral es también el más frecuente decrece conforme θ aumenta, de manera que para $\theta > 5$ esta probabilidad se encuentra por debajo del 50%.

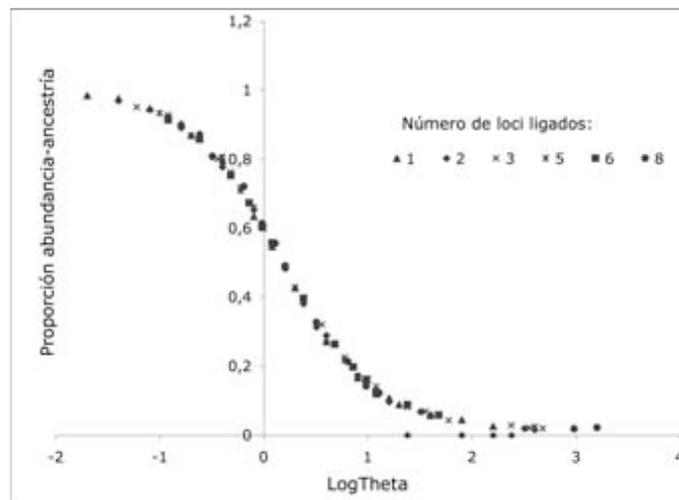


Figura 4.17 Proporción en que el haplotipo más ancestral es también el más frecuente para microsatélites con 1 a 8 loci ligados, en relación a $\theta (2N_e\mu)$.

4.8.1 *Conectividad y abundancia en redes.* Otro atributo que comúnmente se vincula al haplotipo más ancestral es un número elevado de conexiones en una red de haplotipos. Redes construidas con variantes multilocus tomadas de simulaciones con dos distintos grupos de parámetros: $N_e=50,000$; $\mu=10^{-5}$; $k=250$ y $N_e=300,000$; $\mu=10^{-6}$; $k=250$, respectivamente (Figs. 4.18 y 4.19) muestran un comportamiento errático. En ambos casos, la variante más ancestral, resaltada en negro, adquiere posiciones diversas en las diferentes redes, desde una ubicación central hasta algún extremo terminal. Así mismo, varía el número de conexiones que posee. También se aprecia que las variantes más abundantes coinciden con poca frecuencia con la más ancestral. En el primer caso (Fig. 4.18), sólo 1/10 veces el haplotipo más abundante fue también el más ancestral (Fig. 4.18, D) y este último

2/10 veces fue aquél con mayor número de conexiones (Fig. 4.18, H e I). Por otra parte, el haplotipo más abundante resultó ser el más conectado un par de veces (Fig. 4.18, F y J). En ninguna ocasión el haplotipo más ancestral reunió las dos condiciones: abundancia y alta conectividad.

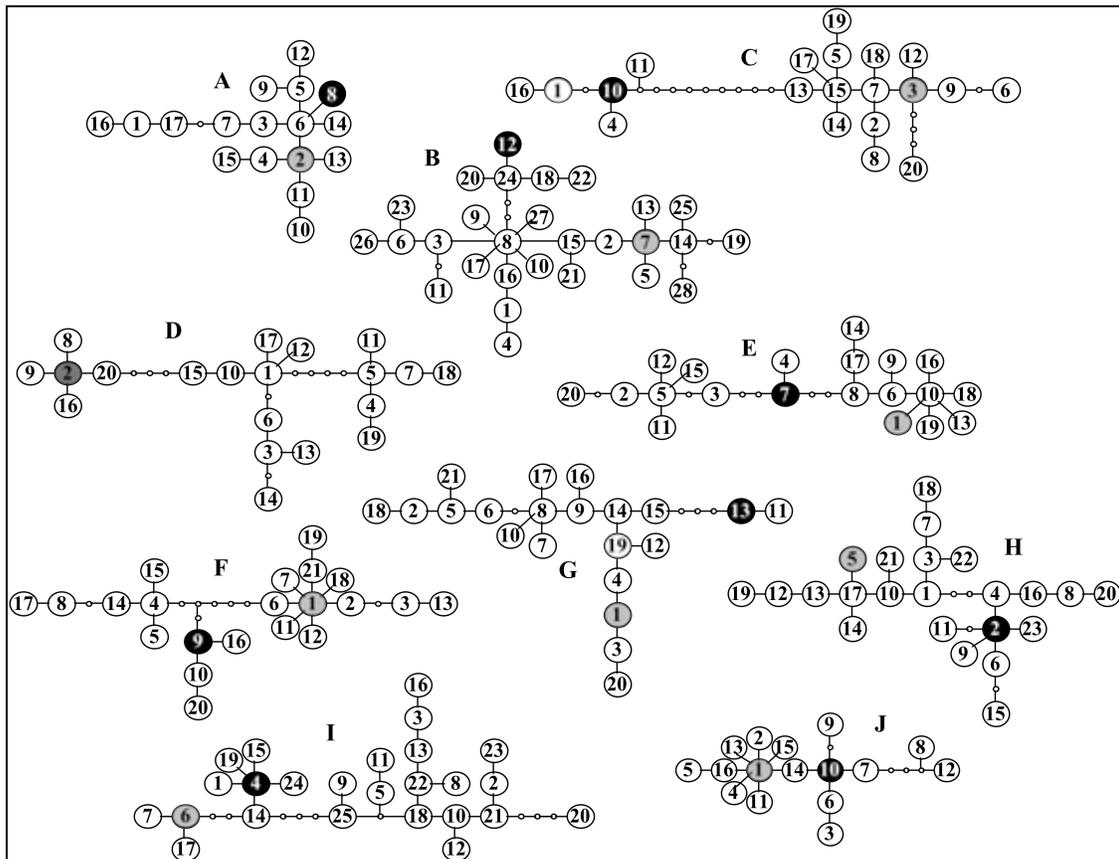


Figura 4.18 Redes de variantes multilocus construidas a partir de datos generados en simulaciones con los parámetros $N_e=50000$, $\mu=10^{-5}$ y $k=250$. La variante más ancestral señalada en negro, la más frecuente en gris claro y cuando ambas condiciones coinciden, en gris oscuro.

Reduciendo la tasa de mutación e incrementando considerablemente el tamaño efectivo poblacional (Fig. 4.19) con el fin de obtener aproximadamente el mismo número de variantes que en el caso anterior, se generan redes más compactas. El haplotipo más ancestral es también el más abundante con una frecuencia de 2/10 (Fig. 4.19, G y J) y posee, o comparte con otra variante, el mayor número de conexiones 5/10 veces (Fig. 4.19, A, B, E, G y H); no obstante, 2 de estas 5 ocasiones el programa TCS V. 1.21 (Clement *et al.*, 2000) reconoce como más ancestral a la variante que comparte el mismo número de

conexiones (Fig. 4.19, E y H, variantes 4 y 5 respectivamente). Respecto a la variante más frecuente, 2/10 veces es también la más conectada (Fig. 4.19, F y G), y una de ellas (Fig. 4.19, F) coincide, a su vez, con la más ancestral.

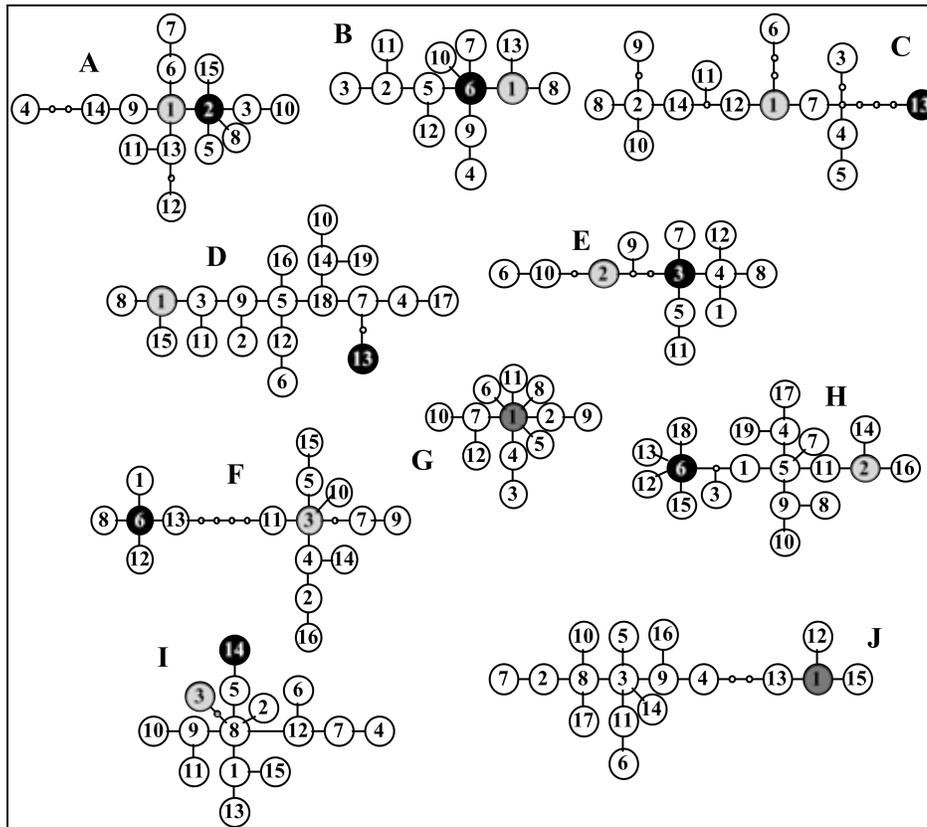


Figura 4.19 Redes de variantes multilocus construidas a partir de datos generados en simulaciones con los parámetros $N_e = 300000$, $\mu = 10^{-6}$ y $k = 250$. La variante más ancestral señalada en negro, la más frecuente en gris claro y cuando ambas condiciones coinciden, en gris oscuro.

4.8.2. *Efecto de la homoplasia sobre la estructura de las redes.* Mediante una primera comparación cualitativa de las redes construidas a partir de dos configuraciones de muestra, variantes multilocus y haplotipos, es decir, con homoplasia estructural y sin ella respectivamente, es evidente que la homoplasia altera la topología de las redes simplificando su estructura (Fig. 4.20; apéndice). Las ramas de las redes construidas a partir de haplotipos se colapsaron cuando en lugar de estos se emplearon variantes multilocus.

Dado lo anterior, se evaluaron ciertas medidas básicas pertenecientes a la teoría de redes, que pudieran describir cuantitativamente las diferencias estructurales entre las redes de haplotipos y las redes de variantes multilocus (cuadro 4.1):

- i. *Tipo de red*: Una red carente de ciclos, que presenta el número mínimo de conexiones para que todos sus nodos permanezcan conectados y en consecuencia, donde cualquier par de nodos está conectado por un único camino, se denomina *árbol*. Las redes de haplotipos y de variantes multilocus pertenecen a esta categoría. Los nodos cuyo grado (numero de conexiones) es 1 se denominan *hojas* (Diestel, 2000).
- ii. *Número de nodos*: El número de nodos (también denominado *orden*) es mayor en las redes de haplotipos que en las redes de variantes multilocus. Cabe recordar en este punto que las simulaciones muestran que la diversidad genética, expresada como el número de haplotipos o variantes distintas, es mayor precisamente en la configuración de la muestra que encierra a los haplotipos. Más aún, las redes de haplotipos no sólo muestran un mayor número de nodos correspondientes a aquellos alelos presentes en la muestra, sino que también exponen un número más elevado de haplotipos no muestreados.
- iii. *Número de hojas*: El número de *hojas* en cada red brinda una aproximación sobre su estructura, de manera que entre más *hojas* posea, su forma será más parecida a una estrella. Por otra parte, entre menos *hojas* presente, será una red más alargada (Diestel, 2000). Las redes de haplotipos presentan mayor número de *hojas* en comparación con las de variantes multilocus, no obstante también presentan mayor número de nodos internos (en su mayoría no muestreados) y en consecuencia, a pesar de que la cantidad de *hojas* podría sugerir que se trata de redes con mayor tendencia a presentar forma de estrella con respecto a las multilocus, el mayor número de nodos internos alarga la estructura general.

- iv. *Diámetro*: La distancia más grande entre dos nodos corresponde al diámetro de la red (Diestel, 2000). Una vez más, las redes de haplotipos en general poseen diámetros mayores en comparación con las de variantes multilocus. A pesar de que las redes de haplotipos poseen mayor número de *hojas*, el número de haplotipos no muestreados, los cuales corresponden a nodos internos, alargan su estructura. Así pues, los nodos que se pierden en las redes de variantes multilocus corresponden tanto a *hojas* como a nodos internos.

Cuadro 4.1. Medidas de las redes de haplotipos y de variantes multilocus.

Red*	No. loci ligados	θ Xu y Fu	SASH	Número de nodos		Número de hojas		Diámetro	
				H ^a	V.M. ^b	H ^a	V.M. ^b	H ^a	V.M. ^b
1	2	1.0739	0.299	9	4	7	3	3	2
2	2	4.1616	0.01875	11	6	5	4	6	3
3	2	1.4271	0.00819	8	5	3	2	6	4
4	2	0.7827	0.0177	10	7	8	5	3	3
5	2	5.1926	0.2238	19	5	11	3	7	3
6	2	14.808	0.498	41	13	14	4	17	8
7	2	11.977	0.2137	28	9	17	4	9	5
8	2	5.4328	0.3616	23	10	12	6	8	5
9	3	6.0132	0.0124	17	11	10	4	7	7
10	3	1.7928	0.0835	28	10	15	7	13	4
11	3	2.4103	0.3095	20	12	11	7	9	6
12	3	1.7951	0.0054	10	7	7	5	4	3
13	6	10.416	0.2746	22	18	10	8	14	11
14	6	49.312	0.0655	45	33	21	17	10	9
15	6	11.014	0.0044	38	23	13	8	16	11
16	6	21.143	0.0087	39	19	14	9	23	10
17	6	30.4004	0.0042	18	16	12	9	5	5

* Las redes se muestran en el apéndice; ^a Haplotipos; ^b Variantes Multilocus

En lo que respecta al haplotipo o variante más ancestral de la muestra, éste es también el más conectado con poca probabilidad, sin embargo, a pesar de que el haplotipo más ancestral puede no ser el más conectado, por lo general ocupa una posición fundamental en las redes al unir dos conjuntos de haplotipos caracterizados por familias de mutaciones distintas (Fig. 4.20). En otras palabras, el haplotipo más ancestral se puede apreciar como un punto a partir del cual se originaron haplotipos que posteriormente se

ramificaron más que éste. En las redes multilocus, correspondientes a las redes de haplotipos que presentan el caso descrito, el haplotipo más ancestral por lo general se ubica en alguna punta de la red (Fig. 4.20).

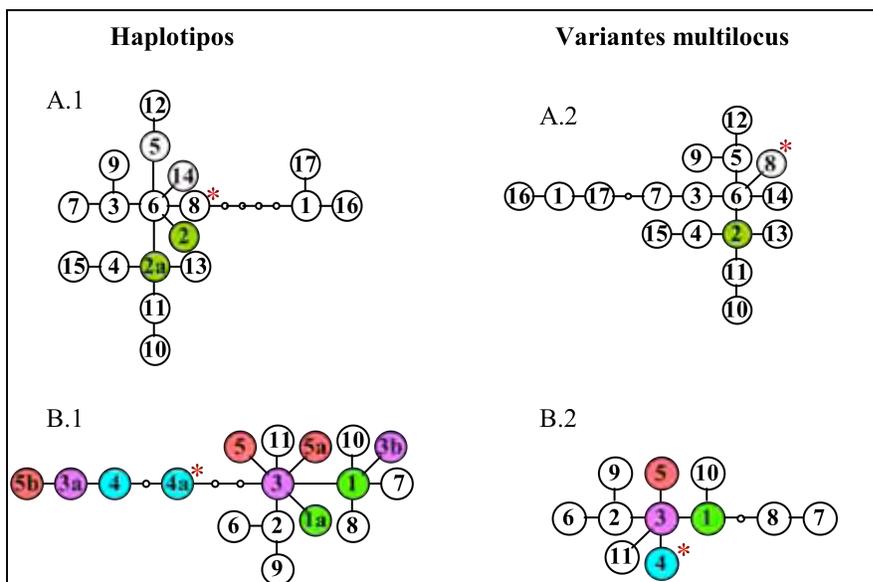


Figura 4.20 Redes de haplotipos (A.1 y B.1) y redes de variantes multilocus (A.2 y B.2), construidas con los siguientes parámetros: A.1 y A.2: $\mu=10^{-5}$, $N_e=50000$, $k=250$, ancestro común= T(11)-T(8)-A(8)-G(9)-T(15)-T(13); B.1 y B.2: $\mu=10^{-5}$, $N_e=50000$, $k=250$, ancestro común= A(8)-G(9)-T(15). Los colores denotan la correspondencia de los haplotipos con su respectiva variante multilocus y el asterisco el haplotipo o variante más ancestral.

Las variantes multilocus homoplásticas por lo general ocupan posiciones centrales dentro de las redes (Fig. 4.20, B.1 y B.2; apéndice), aun cuando los alelos idénticos por secuencia, no por descendencia, se distribuyen en diversas posiciones a lo largo de las redes de haplotipos (Fig. 4.20, A.1 y A.2; apéndice).

V. DISCUSIÓN

5.1 Cálculos respecto a los subconjuntos de datos

De las tres aproximaciones estructurales en relación a las cuales se puede subdividir una muestra, dos de ellas, las variantes por tamaño total o electromorfos y las variantes multilocus, son accesibles mediante métodos experimentales (electroforesis y secuenciación, respectivamente). La heterocigosis y el número de variantes calculados a partir de ambas configuraciones son muy distintos, particularmente cuando el tamaño efectivo poblacional y/o la tasa de mutación son muy elevados y el número de loci ligados es mayor de 2.

La distribución por identidad, posible únicamente mediante simulaciones, ofrece un punto de comparación para los cálculos que se pueden hacer con la información a la que se tiene acceso. Mientras mayor es el número de loci contemplados, las medidas de heterocigosis y el número de variantes generados a partir de la distribución multilocus se ajustan de manera muy precisa a los obtenidos con base en la distribución por identidad.

Lo anterior implica que la utilización de electromorfos subvalora la variación genética (salvo para $\theta < 1$), no obstante el conocimiento de su secuencia, considerando la presencia de loci ligados, aporta información valiosa que contrarresta efectivamente esa subvaloración.

5.2 Homoplasia molecularmente accesible

Otra aplicación significativa derivada de la determinación de los distintos subconjuntos de la muestra, específicamente de la comparación de la heterocigosis calculada para variantes por tamaño total y variantes multilocus, es la definición de la homoplasia molecularmente accesible o *MASH*.

La preocupación generada en torno al problema que significa la presencia de homoplasia en los datos empleados para realizar cálculos genéticos es natural si se considera únicamente la homoplasia por tamaño, *SH*. El cálculo de la homoplasia

molecularmente accesible ofrece un panorama mucho más positivo si se trabaja con loci ligados.

Es cierto que, a causa del mecanismo de mutación por el que los microsatélites evolucionan, se espera una proporción de convergencia por estado sin implicar una relación ancestro-descendiente verdadera, no obstante, la homoplasia molecularmente accesible puede descubrir en una medida importante la cantidad de homoplasia por tamaño. Considerando 3 loci ligados en adelante, la proporción de *MASH* es casi igual a la proporción de *SH*, ambas en función de θ (Fig. 5.1). Incluso tomando en cuenta tan sólo dos loci, lo cual evidentemente es el número mínimo de loci para tener accesibilidad molecular, ya se tiene un aproximación importante a la cantidad de homoplasia total.

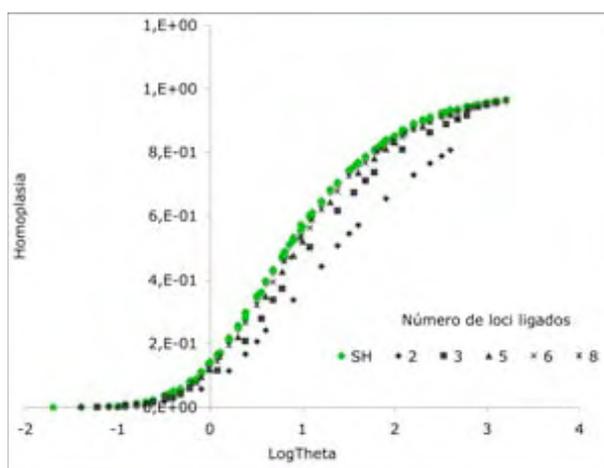


Figura 5.1 Proporción de la homoplasia por tamaño (*SH*; en verde) revelada por la homoplasia molecularmente accesible (*MASH*; en gris), ambas respecto a $\theta (2N_e\mu)$. $k = 200$.

La homoplasia estructuralmente accesible, *SASH* que, al igual que *SH* no se puede calcular directamente, se mantiene como una fracción de homoplasia que *MASH* no puede revelar ya que esta última se fundamenta en el conocimiento de las diferencias entre las secuencias y *SASH* es una medida de la homoplasia que existe entre secuencias aparentemente idénticas pero generadas por mutaciones distintas. La proporción en que *SASH* se presenta decrece conforme más loci se empleen, manteniéndose por debajo de 10% cuando se trabaja con 5 loci ligados o más, en un amplio intervalo de θ de 0.001 a 1000. Para tres loci en un intervalo de θ de 1 a 100 no sobrepasa el 30%, sin embargo, para dos loci rebasa el 50%.

5.3 Homoplasia respecto a θ

El cálculo preciso de θ es determinante, ya que es un parámetro crucial que revela los principales factores que intervienen en la variación genética.

El comportamiento que presentan las diferentes homoplasias en función de θ ($2N_e\mu$) le confiere a este parámetro la particularidad de ser un potencial punto de referencia para teóricamente inferir una proporción de homoplasia una vez que se calcula la θ de una muestra dada, sin embargo, lo anterior depende de la precisión con que θ pueda ser medida. Muchos acercamiento aplicados a microsatélites se han propuesto, calculando θ a partir de la heterocigosis (Ohta y Kimura, 1973), de la varianza entre tamaños de haplotipos (Wehrhahn, 1975), con un enfoque de máxima verosimilitud (Nielsen, 1997) y aplicando una corrección a la θ basada en la heterocigosis por medio de una regresión estadística (Xu y Fu, 2004), entre otros.

Todos los acercamientos tienen sus limitaciones. Por ejemplo, las θ multilocus de Ohta-Kimura (Fig. 5.2, A) y Xu-Fu muestra una sobrevaloración muy significativa respecto a la θ esperada ($\theta = 2N_e\mu$), particularmente para $\theta > 10$ y para 2 o más loci ligados.

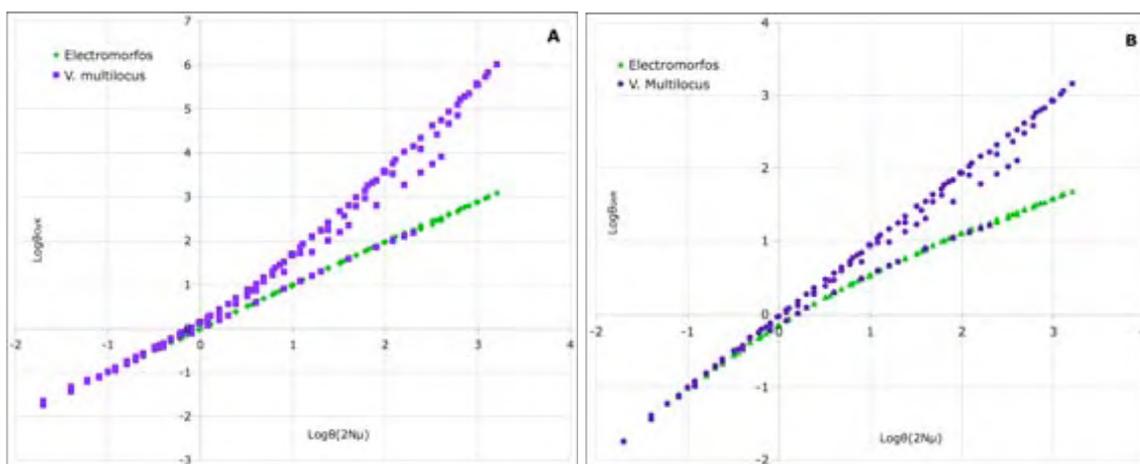


Figura 5.2 Relación entre la θ calculada: A) de acuerdo al modelo SMM (Ohta y Kimura, 1973) y B) de acuerdo al modelo IAM (Kimura y Crown, 1964), respecto a la θ esperada ($2N_e\mu$). Se utilizaron microsatélites de 1 a 8 loci ligados. Al observar las variantes multilocus se aprecia que entre más loci se suman a la estructura, más se eleva el valor de la θ calculada.

La θ_{OyK} calculada con la heterocigosis de los electromorfos guarda una relación muy

cercana con la θ esperada, posiblemente porque los parámetros de Ohta-Kimura y Xu-Fu se modelaron precisamente para electromorfos. El cálculo de θ_{OyK} basado en la heterocigosis depende del modelo de mutación que subyace a la variación genética, ya que ésta no depende únicamente de los parámetros que integran a θ , es decir la tasa de mutación y el tamaño efectivo poblacional, sino que también obedece al patrón de cambio de los alelos. Una visión multilocus sesga el cálculo de θ_{OyK} porque ya no se trata de la variación del tamaño total de un locus paso a paso como describe el modelo bajo el cual se planteó, sino de la variación de un conjunto de fragmentos.

Por lo anterior, se calculó la θ_{IAM} basada en un modelo de alelos infinitos (*IAM*, por sus siglas en inglés; Kimura y Crown, 1964). La figura 5.2 (B) muestra que la θ_{IAM} calculada para los electromorfos evidentemente no es la adecuada para obtener un valor certero de la θ esperada, sin embargo, evidencia que para microsatélites con 2 o más loci, la θ_{IAM} calculada para variantes multilocus es una mejor opción. De manera que cuando se trabaja con variantes multilocus el valor calculado de θ que más se aproxima a su valor teórico esperado es el calculado bajo el modelo de alelos infinitos.

5.4 Modelos de mutación

El modelo de mutación bajo el cual se llevaron a cabo las simulaciones de este estudio corresponde a un modelo simétrico, con una tasa de mutación constante, que describe la ganancia o pérdida de una única unidad microsatelital, independientemente de la longitud inicial del locus. Concretamente, esto lo expone el modelo de mutación por pasos propuesto por Ohta y Kimura (1978). Dicho modelo es el que plantea de manera más sencilla el proceso de mutar de los microsatélites, no obstante, muchos estudios sugieren que los microsatélites evolucionan con una complejidad mayor, por ejemplo ganancia o pérdida de más de una unidad por mutación (Huang *et al.*, 2002), direccionalidad en el proceso de mutación favoreciendo la adición sobre la sustracción (Primmer *et al.*, 1998) o viceversa (Garza *et al.*, 1995; Wierdl *et al.*, 1997) y la dependencia de dicha direccionalidad en función de la longitud del microsatélite (Garza *et al.*, 1995; Xu *et al.*, 2000).

A raíz de estas inconsistencias, se han realizado una serie de modificaciones al modelo de mutación por pasos con el afán de buscar un mejor ajuste a los datos obtenidos

de poblaciones naturales. Esto ha dado origen, por un lado, a una amplia gama de modelos (Di Rienzo *et al.*, 1994; Kruglyak *et al.*, 1998; Calabrese *et al.*, 2001; Sainudiin *et al.*, 2004) y por otro, a la necesidad de explorar dichos modelos a la luz de los resultados que se han obtenido mediante la observación y estudio de poblaciones naturales.

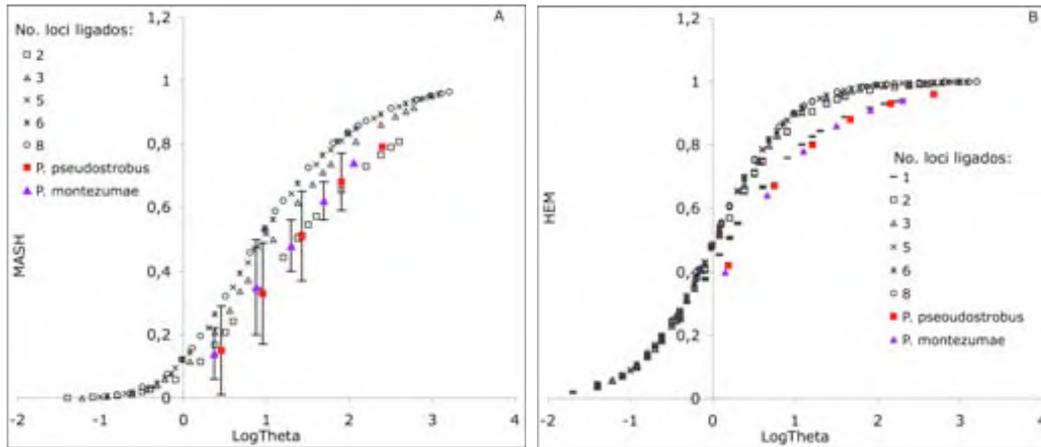


Figura 5.3 Relación entre resultados esperados obtenidos mediante las simulaciones y los calculados para *P. pseudostrobus* y *P. montezumae*: Homoplasia molecularmente accesible (A) y heterocigosis (B).

Con base en lo anterior, a partir de la heterocigosis calculada para dos tipos de datos de las dos especies de pino, *P. montezumae* y *P. pseudostrobus*, se calculó la homoplasia molecularmente accesible para ambas especies y se contrastó con los resultados obtenidos de las simulaciones. La *MASH* obtenida para diferente número de loci (2, 3, 4, 5 y 6) para los pinos muestra una estrecha relación con los resultados obtenidos en las simulaciones (Fig. 5.3A), no obstante, a pesar de que la tendencia es la misma y los valores calculados de *MASH* son similares a los esperados, los cálculos de θ_{XyF} para las dos especies de pino son mayores que lo revelado por las simulaciones, razón suficiente para que el ajuste no sea perfecto. La figura 5.3B muestra, por otra parte, el cálculo de la heterocigosis multilocus tanto para los datos simulados como para los pinos y aquí se aprecia nuevamente la ligera disparidad de la heterocigosis calculada para *P. pseudostrobus* y *P. montezumae* respecto a la heterocigosis arrojada por las simulaciones. Como se discute arriba, el sesgo del cálculo de θ_{XyF} , provocado por un acercamiento multilocus a los datos, puede ser el responsable de la sobrevaloración (Fig. 5.2). De ser así, el modelo de mutación más sencillo propuesto para microsatélites, el cual se probó en este estudio, describiría de una manera muy precisa el

proceso de mutación y las medidas se ajustarían cercanamente. No obstante, cabe la posibilidad de que la disparidad que se observa en los datos sea producto precisamente de la simplicidad del modelo de mutación simulado, de manera que es factible que un modelo más complejo ofrezca un mejor ajuste. La realización de simulaciones que pongan a prueba distintos modelos de mutación es una vía hacia la resolución de esta interrogante.

5.5 Genealogías de haplotipos

5.5.1 Relación ancestro – abundancia. Sería osado aseverar que el haplotipo más frecuente de una muestra corresponde al haplotipo más ancestral de la población o bien, de la muestra misma, particularmente cuando se asocia con poblaciones que arrojan cálculos de θ por arriba de 5. De acuerdo a la teoría de genética de poblaciones, la frecuencia relativa de un haplotipo debe referir directamente a la información sobre su posición ancestral. En otras palabras, la distribución de la probabilidad de las edades de los alelos presentes en un tiempo particular es igual a la distribución de la probabilidad de sus tiempos de extinción, o bien, la probabilidad de que un alelo sea el más antiguo es igual a la probabilidad de que sea el último alelo (de los alelos presentes) en extinguirse debido a mutación y/o deriva génica. Siendo los alelos en cuestión selectivamente neutros, cada uno de ellos tiene igual probabilidad de ser aquel cuyos descendientes sobrevivan durante más tiempo. Así, la probabilidad de que un alelo particular permanezca más tiempo en la población corresponde a su frecuencia relativa. No obstante, cuando se desconocen las frecuencias alélicas de una población, una muestra tomada de la población ofrece ciertas frecuencias y cabe preguntar si el alelo más frecuente en la muestra es el más ancestral de la población (Watterson y Guess, 1977).

A través de las simulaciones se obtuvieron resultados que indican la proporción con que el haplotipo más frecuente de la muestra coincide con el más ancestral y ésta decrece conforme el parámetro θ aumenta, es decir, cuanto más elevada sea la tasa de mutación, el tamaño efectivo poblacional y el número de loci ligados, menor será la probabilidad de que el haplotipo más ancestral sea también el más abundante. Esto es consistente con lo obtenido a través de un enfoque teórico matemático por Watterson y Guess (1977; Fig. 5.4) y por simulaciones realizadas por Ewens y Gillespie en 1974.

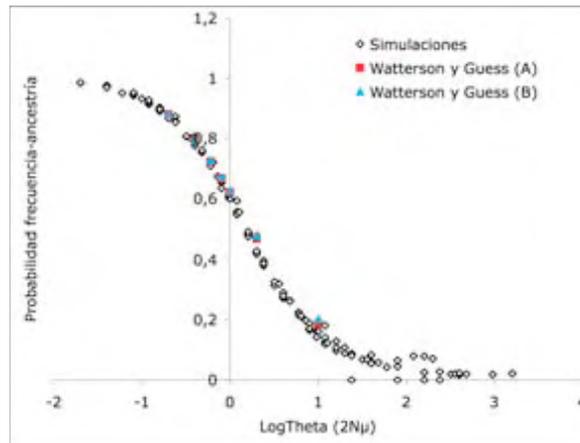


Figura 5.4 Proporción en que el haplotipo más ancestral es también el más frecuente para microsatélites con 1 a 8 loci ligados, en relación a θ ; (A) Probabilidad de que el alelo más frecuente de una muestra de tamaño 100 sea el más ancestral de la población; en rojo; (B) Probabilidad de que el más frecuente de la muestra sea el más ancestral de la muestra; en azul.

Watterson y Guess (1977) inclusive proponen que la probabilidad de que un alelo con una frecuencia relativa n_j/M en una muestra, donde n_j es la frecuencia del alelo j y M el tamaño de la muestra, sea el más ancestral de la población es

$$R_j(n) = n_j / M + \theta, \text{ donde } j = 1, 2, \dots, k.$$

De manera que $n_1 + n_2 + \dots + n_k = M$ y como consecuencia la probabilidad de que el alelo más antiguo se incluya en la muestra es

$$R_1(n) + R_2(n) + \dots + R_k(n) = \sum_1^k n_j / (M + \theta) = M / (M + \theta).$$

El valor de estas expresiones se presenta en la figura 5.4, expresado en la forma del valor esperado de $R_j(n)$, es decir $E(R_j(n)) = [M / (M + \theta)] E(n_j / M)$. $E(R_j(n))$ corresponde a la probabilidad de que el alelo más frecuente de una muestra de tamaño M sea el más ancestral de la población. El planteamiento matemático teórico de esta probabilidad, coincide de manera precisa con la medida obtenida de las simulaciones de la proporción con que el alelo más frecuente de una muestra es también el más ancestral, mostrando la consistencia,

por una parte, del modelo simulado y por otra, de los postulados teóricos de Watterson y Guess (1977).

5.5.2 Topología de redes y homoplasia. Para hablar de la topología de las redes es necesario retomar lo planteado en los resultados respecto a las distribuciones de la muestra, es decir, las diferentes configuraciones que ésta puede adoptar de acuerdo al criterio con que se agrupan las variantes o haplotipos y las definiciones de homoplasia que de ellas derivan.

Mutaciones distintas pueden conducir a secuencias idénticas, oscureciendo la búsqueda de una relación genealógica certera. No obstante, ya que cada mutación no genera una variante única, sino que existe la posibilidad de que ocurra una convergencia de estados por vías de mutaciones diferentes, alelos distintos por descendencia se agrupan en el mismo nodo de una red de variantes multilocus lo cual implica una pérdida de estructura. La homoplasia estructuralmente accesible (*SASH*), describe la relación entre los haplotipos por identidad y las variantes multilocus, al comparar las heterocigosis entre ambos subconjuntos de la muestra. El propósito de construir redes a partir de haplotipos y de sus correspondientes variantes multilocus fue el de estudiar el efecto topológico que la homoplasia estructuralmente accesible (*SASH*) pudiera conferir.

Como se mostró en los resultados, el número de haplotipos excede al número de variantes multilocus, lo cual altera la estructura de las redes multilocus simplificándolas. Las simulaciones revelan que este fenómeno está vinculado al número de loci ligados del microsatélite con cuyos haplotipos y/o variantes se está trabajando, de manera que entre más loci estén ligados menor será la diferencia entre el número de haplotipos y el número de variantes multilocus ($N_H^I - N_V^M$; Fig. 5.5 A y B). Más aún, las simulaciones muestran una tendencia clara de $N_H^I - N_V^M$ en relación con θ y *SASH*, de manera que mientras estas últimas incrementan la diferencia entre el número de haplotipos y el número de variantes multilocus también aumenta (Fig. 5.5 A y B).

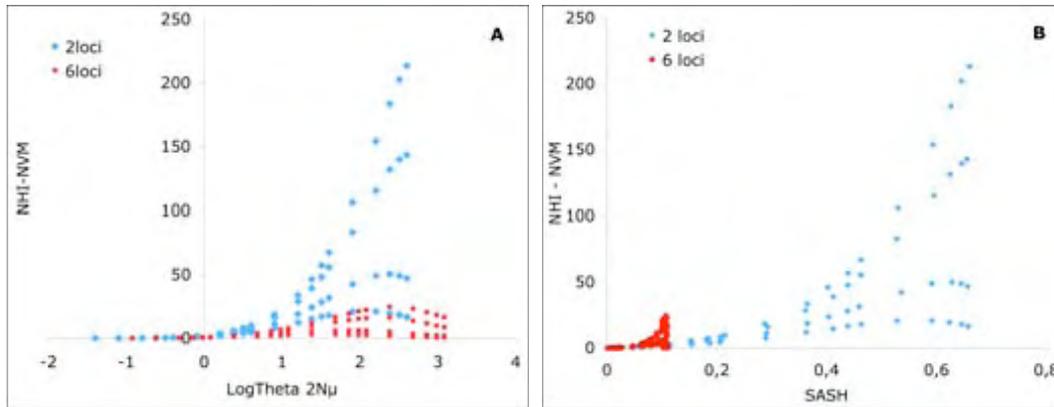


Figura 5.5 Diferencia entre el número de haplotipos y el número de variantes multilocus ($N_H^I - N_V^M$) con respecto a: $\theta = 2N_e\mu$ (A) y a la homoplasia estructuralmente accesible (B). $k = 100, 200, 500$ y 750 ; $s = 2000$.

No obstante, la relación de las medidas del diámetro y del número de *hojas* entre las redes de haplotipos y las redes de variantes multilocus con respecto a la homoplasia estructuralmente accesible falla en mostrar una tendencia evidente. Intuitivamente se esperaría que la diferencia de diámetro y/o número de *hojas* entre las redes siguiera un comportamiento de acuerdo a la proporción de *SASH* que presentaran los datos, de manera que a mayor *SASH*, mayor fuera la diferencia entre las medidas de las redes multilocus y las redes de haplotipos.

El diámetro de las redes puede o no variar a pesar de las diferencias entre el número de haplotipos y el número de variantes, dependiendo de la manera con que se vinculan entre sí los nodos. Esto quiere decir que un mayor número de haplotipos no implica un diámetro mayor, ya que existen nodos muy conectados que pueden reducir la distancia entre el par de nodos más alejados y por lo tanto puede no haber correlación entre el diámetro respecto al número de nodos o *SASH*. Ahora bien, el número de *hojas* es una medida que ofrece una noción sobre la estructura de las redes y podría esperarse una relación entre ésta y *SASH* ya que a primera vista las redes de haplotipos son muy diferentes estructuralmente a las multilocus, sin embargo, los resultados no muestran una relación donde a medida que *SASH* incrementara el número de *hojas* se redujera.

Las simulaciones llevadas a cabo con 2000 duplicados, muestran una tendencia clara entre la diferencia del número de haplotipos y de variantes multilocus ($N_H^I - N_V^M$) respecto a *SASH* (Fig. 5.5 B). Dicha tendencia tampoco se ve al analizar los datos de las

redes, por lo que a pesar de que sus medidas no revelan nada concluyente, las redes son pocas y es posible que al incrementar los datos se obtengan resultados más consistentes.

Cabe enfatizar que los datos con los que se construyeron las redes de este apartado se obtuvieron de simulaciones sencillas donde el efecto de la estocasticidad, inherente al proceso coalescente y al modelo de mutación, es muy elevado. En otras palabras, si se llevan a cabo simulaciones individuales con una combinación de parámetros determinada, cada una de ellas puede generar resultados muy diferentes. De aquí se deriva que el resto de las simulaciones en este estudio se realizara con 2000 duplicados, esto con el afán de reducir la estocasticidad. No obstante generar una muestra de microsátélites para construir una red es posible únicamente mediante una simulación sencilla y por lo tanto es muy posible que el comportamiento que se espera en relación a *SASH* a partir de estas redes no sea concluyente ya que son producto de pocos datos, influenciados por el efecto estocástico.

VI. CONCLUSIONES Y PERSPECTIVAS

1. La variación genética aportada por los microsátelites, particularmente cuando existe ligamiento de dos o más loci polimórficos, es adecuada para realizar estudios de genética de poblaciones.
2. El ligamiento de loci a su vez, es un factor fundamental en la reducción de la homoplasia estructuralmente accesible (*SASH*) ya que la utilización de al menos 3 loci polimórficos la aminora en gran medida.
3. El cálculo de la homoplasia molecularmente accesible (*MASH*) revela una importante proporción de la homoplasia por tamaño (*SH*), aun cuando sólo se emplean dos loci ligados, es decir el número mínimo de loci necesarios para que sea posible su cálculo. A partir de 3 loci *MASH* revela más del 80% de *SH* y de 5 en adelante más de 90%.
4. Ya que no hay un modelo completamente correcto en el sentido de que difícilmente puede abarcar en su totalidad a todos los posibles factores que determinan un cierto comportamiento, sería importante explorar otros modelos de mutación que posiblemente ofrezcan un mejor ajuste a los datos observados.
5. Hacer inferencias en relación a θ parece seguro siempre y cuando su valor calculado en variantes multilocus con base en el modelo SMM, sea muy reducido, preferentemente $\theta < 5$. De lo contrario es posible que los cálculos de θ_{OyK} , y θ_{XyF} sean mayores a lo esperado por la definición $2N_e\mu$. Por ello sería recomendable calcular la θ_{IAM} , cuando se trabaje con variantes multilocus particularmente cuando presentan 3 o más loci ligados, e inclusive cuando sólo cuentan con 2 loci.
6. El postulado que dicta que el haplotipo más frecuente de una muestra es el más ancestral de la población debe emplearse con precaución, particularmente si los datos arrojan cálculos de θ (multilocus) por encima de 5.

7. La homoplasia estructuralmente accesible simplifica la topología de las redes de haplotipos a causa de la subvaloración del número de alelos. Es recomendable la utilización del mayor número de loci polimórficos posible para reducir este fenómeno.

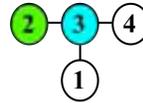
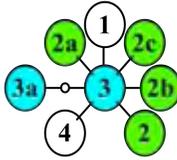
8. Los microsátélites de cloroplasto tienen pues un importante potencial en su papel como marcadores genéticos. Para ello es fundamental entender su dinámica evolutiva y hacia este fin los estudios mediante simulaciones son una eficiente herramienta.

VII. APÉNDICE: Redes^{1,2}

Haplotipos

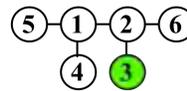
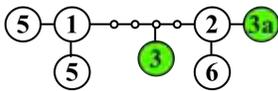
Variantes multilocus

1



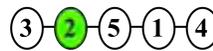
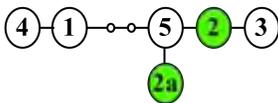
No. Loci = 2
 $\theta = 1.0739$
 SASH = 0.299
 Más ancestral: 2b
 Más frecuente: 3

2



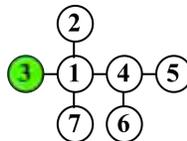
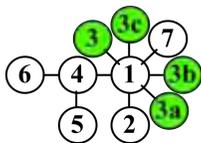
No. Loci = 2
 $\theta = 4.1616$
 SASH = 0.1875
 Más ancestral: 1
 Más frecuente: 1

3



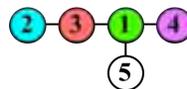
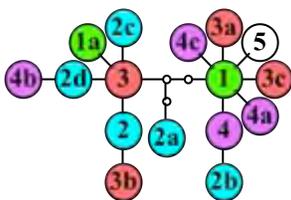
No. Loci = 2
 $\theta = 1.4271$
 SASH = 0.00819
 Más ancestral: 1
 Más frecuente: 1

4



No. Loci = 2
 $\theta = 0.7827$
 SASH = 0.0177
 Más ancestral: 2
 Más frecuente: 1

5



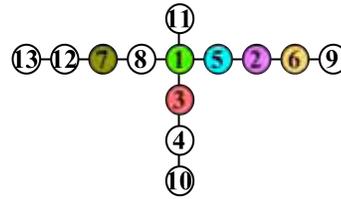
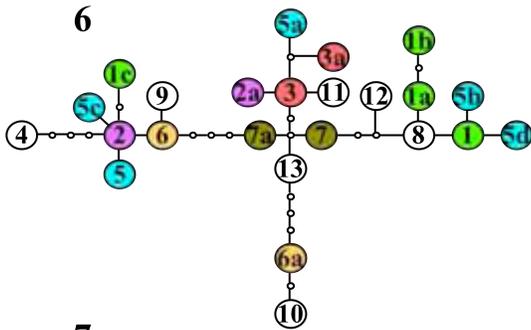
No. Loci = 2
 $\theta = 5.1926$
 SASH = 0.2238
 Más ancestral: 1
 Más frecuente: 1

¹ Los colores denotan los haplotipos homoplásticos y la correspondiente variante multilocus donde se agrupan.

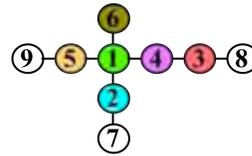
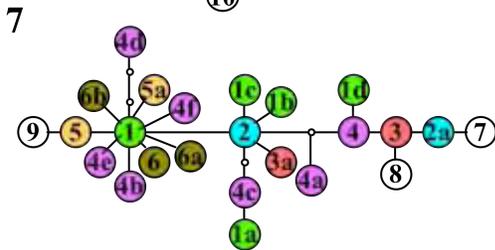
² El tamaño de muestra con que se llevaron a cabo las simulaciones para construir las redes es $k = 250$.

Haplotipos

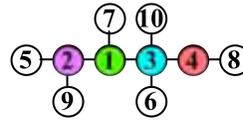
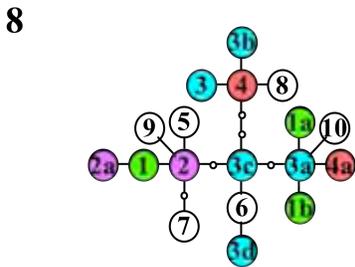
Variantes multilocus



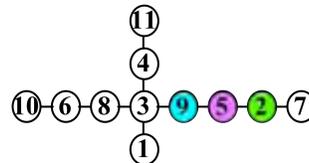
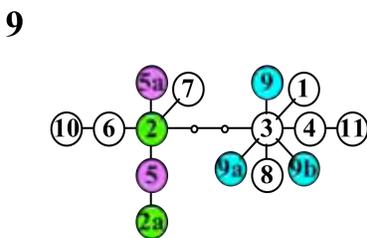
No. Loci = 2
 $\theta = 14.808$
 SASH = 0.498
 Más ancestral: 4
 Más frecuente: 2



No. Loci = 2
 $\theta = 11.977$
 SASH = 0.2137
 Más ancestral: 2
 Más frecuente: 1



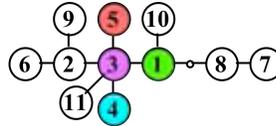
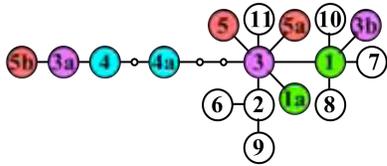
No. Loci = 2
 $\theta = 5.4328$
 SASH = 0.3616
 Más ancestral: 3c
 Más frecuente: 1



Haplotipos

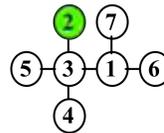
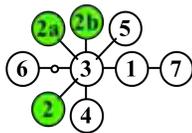
Variantes multilocus

11



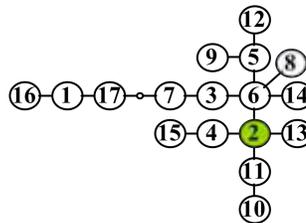
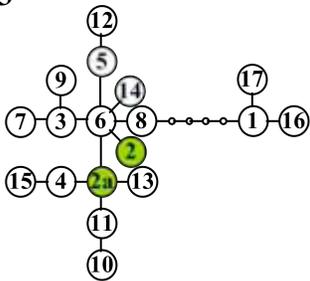
No. Loci = 3
 $\theta = 2.4103$
 SASH = 0.3095
 Más ancestral: 4a
 Más frecuente: 3

12



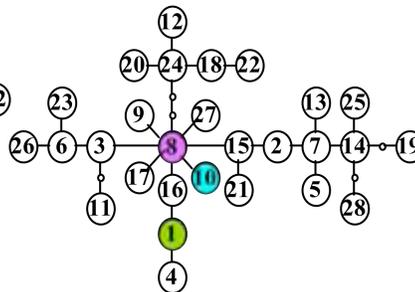
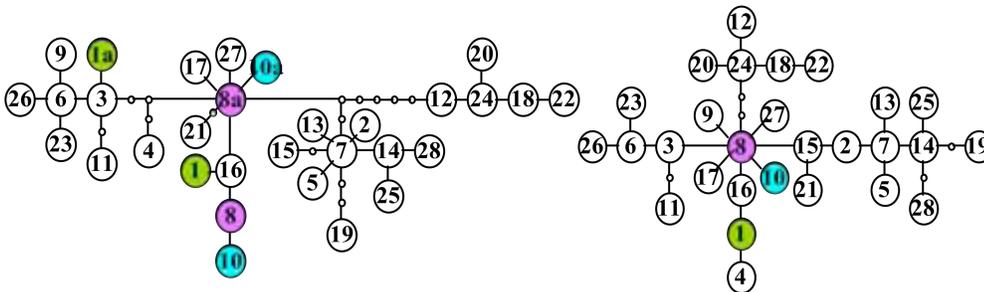
No. Loci = 3
 $\theta = 1.7951$
 SASH = 0.0054
 Más ancestral: 3
 Más frecuente: 1

13



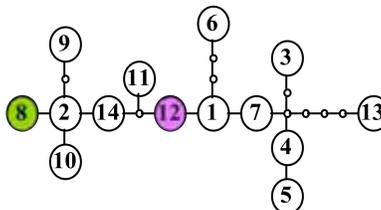
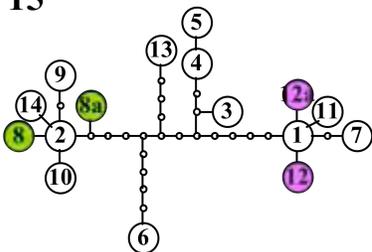
No. Loci = 6
 $\theta = 10.416$
 SASH = 0.2746
 Más ancestral: 8
 Más frecuente: 2

14



No. Loci = 6
 $\theta = 49.312$
 SASH = 0.0655
 Más ancestral: 12
 Más frecuente: 1

15

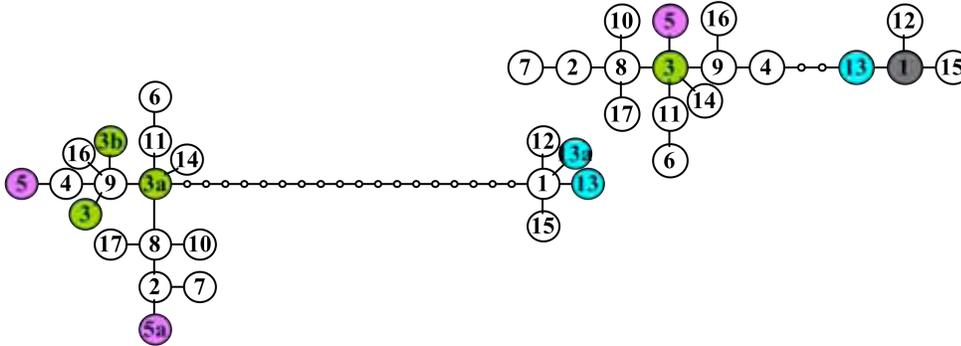


No. Loci = 6
 $\theta = 11.014$
 SASH = 0.0044
 Más ancestral: 13
 Más frecuente: 1

Haplotipos

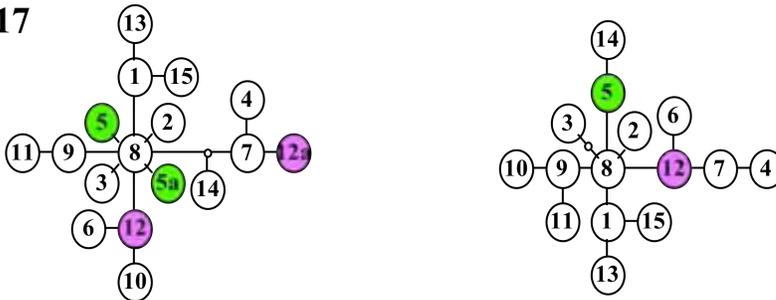
Variantes multilocus

16



No. Loci = 6
 $\theta = 21.143$
 SASH= 0.0087
 Más ancestral: 1
 Más frecuente: 1

17



No. Loci = 6
 $\theta = 30.4004$
 SASH= 0.0042
 Más ancestral: 1
 Más frecuente: 1

VIII. GLOSARIO

Alelo.- Una de las distintas formas que puede adoptar una secuencia de ADN (de un microsatélite para este trabajo) perteneciente a un locus determinado a causa de uno o varios eventos de mutación. En este estudio cada alelo adquiere su identidad al identificar y registrar la o las mutaciones que le dieron origen. Se denomina también *haplotipo*.

Electromorfo.- Fragmento de ADN (microsatélite) caracterizado únicamente por su tamaño y en consecuencia por su movilidad electroforética. También llamado en este trabajo como *variante de tamaño*.

Elementos nucleares dispersos cortos.- SINE o *short interspersed nuclear elements*. Son elementos retrotransponibles no autónomos, dispersos en el genoma, que se prolongan aproximadamente entre 80 a 400 pares de bases. Los SINE más prominentes son los elementos Alu en humanos y los elementos B1 en ratón.

Elementos nucleares dispersos largos.- LINE o *long interspersed nuclear elements*. Son elementos retrotransponibles autónomos, dispersos en el genoma, cuya longitud se extiende hasta secuencias de varias kilobases. Entre los LINE más notables destacan los elementos L1.

Haplotipo.- Combinación de alelos de diferentes loci que están ligados.

Heterocigosis.- Probabilidad de que dos secuencias elegidas al azar correspondan a diferentes tipos alélicos. Se expresa matemáticamente como $He = \frac{n}{n-1} (1 - \sum_i p_i^2)$, donde p es la frecuencia relativa de cada alelo y n es el tamaño de la muestra.

Heterocigosis locus.- Heterocigosis calculada a partir del número y frecuencia de los electromorfos (variantes de tamaño). $H_E^T = \frac{n}{n-1} (1 - \sum_i Vt_i^2)$, donde Vt es la frecuencia relativa de cada variante de tamaño y n es el tamaño de la muestra.

Heterocigosis multilocus.- Heterocigosis calculada a partir del número y frecuencia de las variantes multilocus. $H_E^M = \frac{n}{n-1} (1 - \sum_i Vm_i^2)$, donde Vm es la frecuencia relativa de cada variante multilocus y n es el tamaño de la muestra.

Heterocigosis por identidad.- Heterocigosis calculada a partir del número y frecuencia de los alelos (haplotipos). $H_E^H = \frac{n}{n-1} (1 - \sum_i H_i^2)$, donde H es la frecuencia relativa de cada alelo (haplotipo) y n es el tamaño de la muestra.

Homoplasia.- Fenómeno que ocurre cuando un carácter en dos organismos no es derivado de un carácter correspondiente en su ancestro común más reciente. Para un marcador genético, la homoplasia ocurre cuando diferentes copias de un locus son idénticas en estado pero no son idénticas por descendencia.

Homoplasia de tamaño.- Proporción de variantes de tamaño idénticas por estado pero diferentes por descendencia, es decir, comparten la misma longitud pero varían en secuencia y/o en la serie de mutaciones que las originan. $SH = 1 - [(1 - H_E^I)/(1 - H_E^T)]$

Homoplasia estructuralmente accesible.- Proporción de variantes multilocus idénticas por secuencia pero diferentes por descendencia. $SASH = 1 - [(1 - H_E^I)/(1 - H_E^M)]$

Homoplasia molecularmente accesible.- Fracción de homoplasia por tamaño que es posible discriminar mediante el conocimiento de diferencias en secuencia de los electromorfos. Corresponde a la proporción de variantes homoplásticas estructuralmente accesibles respecto a las homoplásticas de tamaño. $MASH = 1 - [(1 - H_E^M)/(1 - H_E^T)]$

Microsatélite.- Conjuntos de una a ocho pares de bases contiguas repetidas de 10 a 100 veces. Dichos arreglos se prolongan, en promedio, de 20 a 100 pares de bases por locus.

Minisatélite.- Agrupaciones de 10 a 30 pares de bases repetidas hasta conformar bloques de entre 0.5 a 30 kilobases.

Replication slippage.- Error durante la replicación de ADN que conduce al incremento o decremento de unidades repetidas en un arreglo como el de los microsátélites o minisatélites.

Retroelementos.- Elemento genético que se transpone vía un intermediario de RNA.

Retrotransposón.- Elemento ampliamente repetido en el genoma con una secuencia similar a la integrada en un genoma retroviral y posiblemente con actividad de retrotransposición. Se caracterizan, a diferencia de los SINE y LINE, por poseer elementos LTR (*long terminal repeats*) en cada extremo de la secuencia.

Satélite.- ADN ubicado en la heterocromatina de los centrómeros, corresponde a arreglos de 2 a cientos de pares de bases repetidas miles de veces, prolongándose hasta 5 megabases.

Transposón.- Elemento genético con la capacidad de moverse de un lugar a otro dentro de una molécula de ADN.

Variante de tamaño.- ver *electromorfo*.

Variante multilocus.- Variante cuya secuencia es revelada y puede contener en su estructura dos o más loci distinguibles entre si, de manera que variantes idénticas por tamaño total pueden corresponder a secuencias distintas.

IX. REFERENCIAS

1. Adams, R.I., Brown, K.M. y Hamilton, M.B. (2004) The impact of microsatellite electromorph size homoplasy on multilocus population structure estimates in a tropical tree (*Corythophora alta*) and an anadromous fish (*Morone saxatilis*). *Molecular Ecology*, **13**: 2579–2588.
2. Angers, B., Estoup, A. y Jarne, P. (2000) Microsatellite size homoplasy, SSCP, and population structure: A case study in the freshwater snail *Bulinus truncatus*. *Molecular Biology and Evolution*, **17**: 1926-1932.
3. Armour, J.A.L., Alonso-Alegre, S., Miles, S., Williams, L.J. y Badge, R.M. (1999) Minisatellites and mutation processes in tandemly repetitive DNA en Goldstein, D.B. y Schlötterer, C., (Eds.) *Microsatellites: Evolution and Application*. Oxford University Press Inc., New York.
4. Brown, T.A. (2002) *Genomes*. Garland Science. 2 ed. E.U.A.
5. Calabrese, P.P., Durrett, R.T. y Aquadro, C.F. (2001) Dynamics of microsatellite divergence under stepwise mutation and proportional slippage/point mutation model. *Genetics*, **159**: 839-852.
6. Clement, M., Posada, D. y Crandall, K.A. (2000) TCS: a computer program to estimate gene genealogies. *Molecular Ecology*, **9** (10): 1657-1660.
7. Delgado, P., Salas-Lizana, R., Vázquez-Lobo, A., Wegier, A., Anzidei, M., Alvarez-Buylla, E.R., Vendramin, G.G. y Piñero, D. (2007) Introgressive hybridization in *Pinus montezumae* Lamb and *Pinus pseudostrobus* Lindl. (Pinaceae): morphological and molecular (cpSSR) evidence. *International Journal of Plant Sciences*, **168**: 861-875.
8. Diestel, R. (2000) *Graph theory*. Springer-Verlag. 3 ed. E.U.A.
9. Di Rienzo, A., Peterson, A.C., Garza, J.C., Valdes, A.M., Slatkin, M. y Freimer, N.B. (1994) Mutational processes of simple-sequence repeat loci in human populations. *Proc Natl Acad Sci USA*, **91**: 3166 3170.
10. Doyle, J.J., Morgante, M., Tingey, S.V. y Powell, W. (1998) Size homoplasy in chloroplast microsatellites of wild perennial relatives of soybean (*Glycine* subgenus *Glycine*). *Molecular Biology and Evolution*, **15**(2): 215-218.

11. Ellegren, H. (2004) Microsatellites: simple sequences with complex evolution. *Nature Reviews Genetics*, **5**: 435-445.
12. Estoup, A., Jarne, P. y Cornuet, J.M. (2002) Homoplasy and mutation model at microsatellite loci and their consequences for population genetics analysis. *Molecular Ecology*, **11**: 1591–1604.
13. Ewens, W.J. y Gillespie, J.H. (1974) Some simulation results for the neutral allele model with interpretations. *Theoretical Population Biology*, **6**: 35-37.
14. Garza, J. C., Slatkin, M. y Freimer, N.B. (1995) Microsatellite allele frequencies in humans and chimpanzees, with implications for constraints on allele size. *Molecular Biology and Evolution*, **12**: 594-603.
15. Goldstein, D.B. y Clark A.G. (1995) Microsatellite variation in North American populations of *Drosophila melanogaster*. *Nucleic Acids Research*, **23**(19): 3882-3886.
16. Goldstein, D.B. y Schlötterer, C. (1999) *Microsatellites: Evolution and Application*. Oxford University Press Inc., New York.
17. Hancock, J.M. (1999) Microsatellites and other simple sequences: genomic context and mutational mechanisms en Goldstein, D.B. y Schlötterer, C. (Eds.) *Microsatellites: Evolution and Application*. Oxford University Press Inc., New York.
18. Griffiths, A.J.F., Lewontin, R. C., Gelbart, W.M., Suzuki, D.T. y Miller, J.F. (1999) *An introduction to genetic analysis*. W. H. Freeman & Co. 7ed. E.U.A.
19. Hein, J., Schierup, M.H., Wiuf, G. (2005) *Gene Genealogies, Variation and Evolution: A Primer in Coalescent Theory*. Oxford University Press.
20. Heyer, E., Puymirat, J., Dieltjes, P., Bakker, E. y de Knijff, P. (1997) Estimating Y chromosome specific microsatellite mutation frequencies using deep rooting pedigrees. *Human Molecular Genetics*, **6**(5): 799-803.
21. Huang, Q.Y., Xu, F.H., Shen, H., Deng H.Y., Liu, Y.J., Liu, Y.Z., Li, J.L., Recker, R.R. y Deng, H.W. (2002) Mutation Patterns at dinucleotide microsatellite loci in humans. *American Journal of Human Genetics*. **70**: 625-634.

22. Ishii, T., Mori, N. y Ogihara, Y. (2001) Evaluations of allelic diversity at chloroplast microsatellite loci among common wheat and its ancestral species. *Theoretical and Applied Genetics*. **103**: 896-904.
23. Jakobsson, M.S., Säll, T., Lind-Halldén, C. y Halldén, C. (2007) Evolution of chloroplast mononucleotide microsatellites in *Arabidopsis thaliana*. *Theoretical and Applied Genetics*, **114**: 223-235.
24. Kashi, Y. y Soller, M. (1999) Functional roles of microsatellites and minisatellites en Goldstein, D.B. y Schlötterer, C. (Eds.) *Microsatellites: Evolution and Application*. Oxford University Press Inc., New York.
25. Kayser, M., Roewer, L., Hedman, M., Henke, L., Henke, J., Brauer, S., Krüger, C., Krawczak, M., Nagy, M., Dobosz, T., Szibor, R., de Knijff, P., Stoneking, M., y Sajantila, A. (2000) Characteristics and frequency of germline mutations at microsatellite loci from the human Y chromosome, as revealed by direct observation in father/son pairs. *American Journal of Human Genetics*. **66**:1580-1588.
26. Kimura, M. y Crown, J.F. (1964) The number of alleles that can be maintained in a finite population. *Genetics*. **49**: 725-738.
27. Kruglyak, S., Durrett, R.T., Schug, M.D. y Aquadro, C.F. (1998) Equilibrium distributions of microsatellite repeat length resulting from a balance between slippage events and point mutations. *Proceedings of the National Academy of Sciences*. **95**: 10774-10778.
28. Levinson, G. y Gutman, G.A. (1987) High frequencies of short frameshifts in poly-CA/TG tandem repeats borne by bacteriophage M13 in *Escherichia coli* k-12. *Nucleic Acids Research*, **15**: 5323-5338.
29. Majewsky, J. y Ott, J. (2000) GT repeats are associated with recombinations on human chromosome 22. *Genome Research*, **10**: 1108-1114.
30. Marshall, H.D., Newton, C. y Ritland, K. (2002) Chloroplast phylogeography and evolution of highly polymorphic microsatellites in lodgepole pine (*Pinus contorta*). *Theoretical and Applied Genetics*. **104**: 367-378.
31. Mogg, M., Batley, J., Hanley, S., Edwards, D., O'Sullivan, H. y Edwards, K.J. (2002) Characterization of the flanking regions of *Zea mays* microsatellites reveals

- a large number of useful sequence polymorphisms. *Theoretical and Applied Genetics*, **105**:532-542.
32. Morgante, M., Hanafey, M. y Powell, W. (2002) Microsatellites are preferentially associated with nonrepetitive DNA in plant genomes. *Nature Genetics*, **30**: 194-200.
 33. Navascués, M. y Emerson, B.C. (2005) Chloroplast microsatellites: measures of genetic diversity and the effect of homoplasy. *Molecular Ecology*, **14**: 1333–1341.
 34. Navascués, M., Vaxevanidou, Z., González-Martínez, S.C., Climent, J., Gil, L. y Emerson, B.C. (2006) Chloroplast microsatellites reveal colonization and metapopulation dynamics in the Canary Island pine. *Molecular Ecology*, **15**: 2691–2698.
 35. Nei, M. (1987) *Molecular Evolutionary Genetics*. Columbia University Press, New York.
 36. Nielsen, R. (1997) A likelihood approach to population samples of microsatellite alleles. *Genetics*, **146**: 711–716.
 37. Nordborg, M. (2001) *Coalescent theory* en: Balding, D., Bishop, M. y Cannings, C. *Handbook of Statistical Genetics*. Wiley. Chichester, UK, pp. 179-212.
 38. Ohta, T. y Kimura, M. (1973) A model of mutation appropriate to estimate the number of electrophoretically detectable alleles in a finite population. *Genetical Research*, **22**: 201–204.
 39. Ohta, T. y Kimura, M. (1978) Stepwise mutation model and distribution of allelic frequencies in a finite population. *Proceedings of the Nacional Academy of Sciences*, **75**(6): 2868-2872.
 40. Petes, T.D., Greenwell, P.W. y Dominska, M. (1997) Stabilization of microsatellite sequences by variant repeats in yeast *Saccharomyces cerevisiae*. *Genetics*, **146**: 491-498.
 41. Posada, D. y Crandall, K.A. (2001) Intraspecific gene genealogies: trees grafting into networks. *Trends in Ecology and Evolution*, **16**(1): 37-45.
 42. Primmer, C.R., Saino, N. Moller, A.P. y Ellegren, H. (1998) Unraveling the process of microsatellite evolution through analysis of germ line mutations in barn swallows *Hirundo rustica*. *Molecular Biology and Evolution*, **15**(8): 1047-1054.

43. Provan, J., Powell, W. y Hollingsworth, P.M. (2001) Chloroplast microsatellites: new tools for studies in plant ecology and evolution. *Trends in Ecology and Evolution*, **16**(3): 142-147.
44. Provan, J., Soranzo, N., Wilson, N.J., Goldstein, D.B. y Powell, W. (1999) A low mutation rate for chloroplast microsatellites. *Genetics*, **153**: 943-947.
45. Ribeiro, M.M., Plomion, C., Petit, R., Vendramin, G.G. y Szmidt, A.E. (2001) Variation in chloroplast single-sequence repeats in Portuguese maritime pine (*Pinus pinaster* Ait.). *Theoretical and Applied Genetics*, **102**: 97-103.
46. Rose, O. y Falush, D. (1998) A threshold size for microsatellite expansion. *Molecular Biology and Evolution*, **15**(5): 613-615.
47. Rubinsztein, D.C. (1999) Trinucleotide expansion mutations cause diseases which do not conform to classical Mendelian expectations en Goldstein, D.B. y Schlotterer, C. (Eds.) *Microsatellites: Evolution and Application*. Oxford University Press Inc., New York.
48. Sibly, R.M., Meade, A., Boxall, N., Wilkinson, M.J., Corne, D.W. y Whittaker, J.C. (2003) The structure of interrupted human AC microsatellites. *Molecular Biology and Evolution*, **20**: 453-459.
49. Sainudiin, R., Durrett, R.T., Aquadro, C.F. y Nielsen, R. (2004) Microsatellite mutation models: insights from a comparison of humans and chimpanzees. *Genetics*, **168**: 383-395.
50. Treco, D. y Arnheim, N. (1986) The evolutionarily conserved repetitive sequence d(TG · AC)_n promotes reciprocal exchange and generates unusual recombinant tetrads during gene meiosis. *Molecular and Cellular Biology*, **6**(11): 3934-3947.
51. van Oppen, M. J. H., Rico, C., Turner, J.F., y Hewitt, G.M. (2000) Extensive homoplasy, nonstepwise mutations, and shared ancestral polymorphism at a complex microsatellite locus in Lake Malawi cichlids. *Molecular Biology and Evolution*, **17**(4): 489-498.
52. Watterson, G.A. y Guess, H.A. (1977) Is the most frequent allele the oldest?. *Theoretical Population Biology*, **11**: 141-160.
53. Wehrhahn, C.F. (1975) The evolution of selectively similar electrophoretically detectable alleles in finite natural populations. *Genetics*, **80**: 375-394.

54. Wierdl, M., Dominska, M. y Petes, T.D. (1997) Microsatellite instability in yeast: dependence on the length of the microsatellite. *Genetics*, **146**: 769–779.
55. Wimberger, D. y Piñero, D. (2007) *Gye-coalescence (Coalescence simulation and analysis framework)*. V. 0.1. [Programa informático en línea]. Disponible en Internet en: <<http://code.google.com/p/gye-coalescence>> [con acceso el 2 de septiembre de 2007).
56. Xu, H. y Fu, Y. X. (2004) Estimating effective population size or mutation rate with microsatellites. *Genetics*, **166**: 555-563.
57. Xu, X., Peng, M. y Fang, Z. (2000) The direction of microsatellite mutations is dependent upon allele length. *Nature Genetics*, **24**: 396-399.