

5. Coeficientes de endogamia y estructura poblacional

Enrique Lessa

2024-10-07

Introducción

La finalidad de este documento es presentar, a un nivel introductorio, una visión general del problema de la estructura poblacional y el uso de coeficientes de endogamia (también llamados de endocría) para su estudio. En particular, se destaca una perspectiva genealógica, que usa las propiedades medias de pares de alelos para describir la estructura poblacional. El énfasis está en los conceptos y las fórmulas que los reflejan, sin preocuparse por ajustes para muestras empíricas o modelos mutacionales. Éstos son obviamente importantes para las aplicaciones, pero no para esta introducción.

Los coeficientes de endogamia (estadísticos F) fueron desarrollados por *Sewall Wright (1951)* en base a modelos de estructura poblacional y descripciones generales de la variación genética (frecuencias alélicas y su varianza a distintos niveles de organización de las poblaciones). La correspondencia de estos estadísticos con las genealogías de alelos fue formalizada mucho después por *Montgomery Slatkin (1991)* en base a la teoría del coalescente (*Kingman 1982*). En este documento partimos de la descripción genealógica y demostramos (a veces; en otros casos simplemente enunciamos) su correspondencia con varios estimadores, incluyendo los estadísticos F de Wright.

Se ha optado por presentar estos estadísticos en primer lugar como descriptores de la variación genética, independientemente de los factores y procesos (e.g., patrones de apareamiento locales, estructura geográfica, selección natural) que pueden subyacer a los valores observados. Lamentablemente, los coeficientes de endogamia no nulos no siempre resultan de la endogamia, de modo que el nombre, que tiene raíces históricas, no es muy afortunado para toda la familia de estadísticos F .

Luego de presentar los estadísticos F como descriptores mostraremos su comportamiento bajo ciertos modelos sencillos, incluyendo algunos casos de genes bajo selección natural y algunos modelos de estructura poblacional.

1. Estructura poblacional

Consideremos primero una situación por demás común en biología: los organismos de una misma especie parecen estar organizados de modo de formar subunidades dentro de una unidad mayor. Así, por ejemplo, los líquenes forman parches discretos sobre superficies rocosas, las cuales, además, tampoco tienen una distribución uniforme, sino que se encuentran en distintos afloramientos de rocas, separados por áreas no rocosas. Del mismo modo, los chimpancés están organizados en bandas, grupos sociales reconocibles dentro de un bosque; a su vez, los bosques están fragmentados en el espacio de distintas maneras.

Podríamos suponer que estas unidades poblacionales se han diferenciado genéticamente, al menos en cierta medida, pero no podemos descartar que, si bien los organismos se asocian entre sí y/o con el ambiente en unidades discretas, todas ellas

sean partes de una gran población genéticamente uniforme. Por ejemplo, los grupos sociales podrían ser temporales y rearmarse con frecuencia, o que el régimen de apareamiento sea independiente de ellos.

1.1. Enfoque genealógico

Brevemente, el enfoque genealógico, representado por la teoría del coalescente, estudia la ancestría de los alelos de las poblaciones y las mutaciones a ella asociadas. Para el propósito de este documento, será suficiente considerar la ancestría de pares de alelos. Cada par consiste en dos copias de un mismo gen muestreadas en nuestras poblaciones de interés (veremos más adelante cómo diseñamos el muestreo). En el caso más sencillo, que será suficiente para nuestros propósitos, no hay recombinación intragénica, de modo que los alelos tienen una única línea de ascendencia. En ausencia de recombinación, la historia de un par de alelos es muy sencilla, y consiste en:

- a) genealogía: los dos alelos derivan de un ancestro común, de antigüedad t (tiempo medido en generaciones); por tanto, cada alelo está unido al ancestro por una rama de longitud t ;
- b) variación genética: recordemos que cada rama consiste en t generaciones; a lo largo de una rama, un alelo de una generación cualquiera t_i es una copia de su ancestro en la generación precedente t_{i-1} . Cada evento de transmisión es una oportunidad de mutación, es decir de que la secuencia del alelo descendiente adquiera alguna diferencia con la de su ancestro, y contamos con $2t$ generaciones para que esto ocurra.

Aplicaremos el enfoque genealógico al estudio de la estructura poblacional del siguiente modo:

- a) Consideramos (al menos) dos niveles jerárquicos de organización, uno anidado dentro del otro. Por ejemplo, individuos dentro de una población local, parches de hábitat dentro de una región ocupada por nuestra especie de interés, grupos sociales en una población de primates, etc.
- b) Etiquetamos al nivel inferior con la letra X y al superior con la letra Y y consideramos la ancestría promedio de pares de alelos muestreados al azar en cada uno de los dos niveles. Tendremos k subunidades tales que $X_1, X_2, \dots, X_k \in Y$.

La hipótesis nula es que la población es genéticamente homogénea, de modo que las unidades del nivel inferior X son idénticas entre sí porque no son más que réplicas de lo que ocurre en el nivel superior Y . Por tanto, $t_X = t_Y$. Alternativamente, las poblaciones pueden estar estructuradas de muy distintos modos. En términos genealógicos, podemos describir los apartamientos de la hipótesis nula con el siguiente estadístico:

$$t_{XY} = \frac{\overline{t_Y} - \overline{t_X}}{\overline{t_Y}}$$

[1]

El numerador es simplemente la diferencia entre la antigüedad promedio de pares de alelos tomados en el nivel superior Y y el inferior X . Dicha diferencia está normalizada al dividirla por \bar{t}_Y . La normalización determina que t_{XY} varíe entre -1 y 1.

La **Fig. 1** representa gráficamente los conceptos enunciados hasta ahora.

Observamos que, bajo la hipótesis nula (la población Y es genéticamente homogénea), $t_{XY} = 0$. Si los alelos dentro de las subunidades X están más emparentados entre sí que alelos tomados al azar de toda la población Y , el numerador es positivo y, en consecuencia $t_{XY} > 0$. Si, por el contrario, los alelos dentro de las unidades X están, en promedio, más emparentados entre sí que pares de alelos tomados al azar, entonces $t_{XY} < 0$. Estas dos situaciones son bastante comunes. En cambio, veremos que $t_{XY} < 0$ puede ocurrir en circunstancias más restringidas.

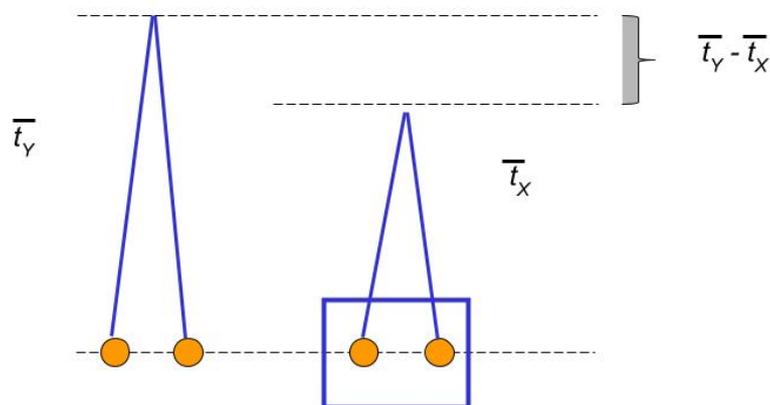


Fig 1. Antigüedad media del ancestro común de dos pares de alelos tomados en un nivel superior (Y) y en un nivel inferior (X) anidado en el anterior.

1.2. Genealogías y variación genética.

Normalmente es imposible saber el tiempo de divergencia de pares de alelos. Por lo tanto, usaremos datos genéticos como indicadores (“proxies”) de dicha divergencia. Ya mencionamos que cada evento de transmisión de un alelo (de su ancestro a su descendiente en la generación siguiente) es una oportunidad de mutación. Consideramos ahora dos modelos de mutación sencillos, en los cuales, como la probabilidad de mutación (por alelo y por generación es muy baja), no hay homoplasia y al comparar dos alelos en el presente vemos todos los cambios ocurridos a lo largo de las dos ramas que los conectan a su ancestro. En particular:

Modelo de alelos infinitos (que debería llamarse modelo de clases alélicas infinitas): El número de clases alélicas es tan grande y la probabilidad de mutación es tan baja que

cada una aparece una única vez en la población. La probabilidad de que dos alelos sean diferentes (pertenecan a clases distintas) H es, en este contexto, proporcional a la antigüedad de su ancestro común:

$$H_X \propto \bar{t}_X, \text{ y } H_Y \propto \bar{t}_Y.$$

Definimos

$$F_{XY} = \frac{H_Y - H_X}{H_Y} \quad [2]$$

y, como consecuencia de lo antedicho, $F_{XY} = t_{XY}$.

En palabras, F_{XY} es la diferencia entre la heterocigosidad media del nivel superior Y y la heterocigosidad media del nivel inferior X , normalizada al dividirla por la primera. Recordemos que la heterocigosidad es la probabilidad de que dos alelos tomados al azar sean distintos (pertenecan a distintas clases alélicas). Comúnmente, la estimamos a partir de las frecuencias alélicas como:

$$\hat{H} = \sum_{i < j}^k 2p_i p_j$$

Si hay varias subunidades y queremos un valor general de F_{XY} , usaremos \bar{H}_X , el promedio de $H_{X_1}, H_{X_2}, \dots, H_{X_k}$, calculados en base a las frecuencias alélicas en cada subunidad. H_Y se obtiene en base a las frecuencias alélicas del nivel superior Y (típicamente estimada como el promedio de las frecuencias alélicas de las distintas subunidades).

Modelo de sitios infinitos El número de sitios a lo largo de un gen es tan grande y la probabilidad de mutación es tan baja que cada sitio muta, como máximo, una única vez. El número de diferencias observadas entre dos alelos es proporcional a la antigüedad de su ancestro común. Llamamos π al número promedio de diferencias entre pares de alelos, con lo cual:

$$\pi_X \propto \bar{t}_X, \pi_Y \propto \bar{t}_Y.$$

Usando π en lugar de H como estimador de t , obtenemos:

$$F_{XY} = \frac{\pi_Y - \pi_X}{\pi_Y} \quad [3]$$

El procedimiento de estimación es semejante al antes descrito para el modelo de alelos infinitos, pero aquí no nos limitamos a considerar las clases alélicas, sino que consideramos también el número de diferencias entre las secuencias que las caracterizan:

$$\hat{\pi} = \sum_{i < j}^k 2p_i p_j \hat{\pi}_{ij}$$

[4]

Observamos que la ecuación [3] es idéntica a la presentada más arriba para \hat{H} [2], excepto que agregamos, al usar [3], el factor π_{ij} , el número de diferencias observadas entre los alelos de las clases i y j . [2] es un caso particular de [3] en el cual $\pi_{ij} = 1$. Si no conocemos las diferencias entre distintas clases alélicas o decidimos ignorarlas, las estamos aproximando a su valor mínimo (1); en efecto, dos alelos de distinta clase tienen al menos una diferencia.

Vamos a hacer una única excepción aquí a lo anunciado al comienzo en cuanto a que vamos a ignorar, en toda nuestra discusión, ajustes de las fórmulas para estimadores empíricos. En el caso de $\hat{\pi}$, el cálculo usual incluye un factor de corrección para el tamaño muestral para obtener un estimador no sesgado. La fórmula aplicada es:

$$\hat{\pi} = \frac{n}{n-1} \sum_{i < j}^k 2p_i p_j \hat{\pi}_{ij}$$

[5]

Notamos que el factor de corrección $\frac{n}{n-1}$ tiende a 1 cuando n tiende a infinito. Para muestras muy grandes, podemos ignorarlo, que es lo que hacemos en general a lo largo de todo este documento.

Modelos con homoplasia No los desarrollaremos en este documento, pero existen muchos casos en los que los genes de interés tienen tasas de mutación grandes y la homoplasia está presente, aún dentro de una población. Un ejemplo común son los loci de microsatélites, para los que es bastante probable que alelos de una misma clase (por ejemplo, un alelo que llamamos CT50 porque tiene 50 motivos de repetición CT) hayan aparecido más de una vez. En otras palabras, los alelos que son idénticos en estado (tienen igual secuencia) no son copias de un único alelo ancestral (no son necesariamente idénticos por descendencia). La clase alélica CT50 es un agregado de alelos que llegaron a ese estado por distintos caminos mutacionales.

En estos casos, necesitamos un modelo mutacional que nos permita estimar distancias genéticas entre alelos que, en promedio, sean proporcionales a sus tiempos de divergencia.

Ejercicio 1: variación en subpoblaciones

Bell et al. (2021) estudiaron la variación geográfica en una especie de ardilla (*Tamias amoenus*) en el suroeste de Estados Unidos, genotipando 11 loci de microsatélites y secuenciando el gen del citocromo b mitocondrial en individuos de varias subpoblaciones. Uno de los objetivos era establecer si la variación genética en la subpoblación de Pine Forest era comparable a la de otras estudiadas o estaba reducida.

Para este ejercicio, seleccionamos un subconjunto de 20 individuos (10 de Pine Forest y 10 de la subpoblación de Warner Mts.) tomados al azar de las secuencias generadas en

el trabajo citado. De los 701 pares de bases reportados en el artículo, se seleccionaron 620 sitios, luego de excluir sitios no secuenciados en todos los individuos. De ellos, 50 fueron polimórficos. Notamos que el número total de sitios polimórficos en las secuencias es aproximadamente el 8%.

En la tabla de resumen, cada tipo de secuencia (haplotipo) es una clase alélica:

		Pine forest	Warner Mts.	Total
Haplo_1_00	<u>GGCCTTACCTCCTTACCTCTCCCCCTCCTTTTTTCGCATCCTACTCTTC</u>	1	5	6
Haplo_2_94	<u>.....C.....</u>	7		7
Haplo_3_06	<u>.....C.....A.....T.....</u>	2		2
Haplo_4_14	<u>AAT.CCGTTCTTAC..TC.CTTTTT.TTCCC.CCTA.TCTTC..CT.CT</u>		3	3
Haplo_5_73	<u>AATTCGGTTCTTACGTTCTCTTTTT.TTCCCCCTA.TCTTC.TCTCCT</u>		1	1
Haplo_6_74	<u>AATTCGGTTCTTACGTTCTC.TTTTTCTCCCCC.ATTCTCGTCTCCT</u>		1	1
		Total 10	10	20
Pine forest				
Haplo_1_00	<u>GGCCTTACCTCCTTACCTCTCCCCCTCCTTTTTTCGCATCCTACTCTTC</u>	1		
Haplo_2_94	<u>.....C.....</u>	7		
Haplo_3_06	<u>.....C.....A.....T.....</u>	2		
		Total 10		
Warner Mts.				
Haplo_1_00	<u>GGCCTTACCTCCTTACCTCTCCCCCTCCTTTTTTCGCATCCTACTCTTC</u>		5	
Haplo_4_14	<u>AAT.CCGTTCTTAC..TC.CTTTTT.TTCCC.CCTA.TCTTC..CT.CT</u>		3	
Haplo_5_73	<u>AATTCGGTTCTTACGTTCTCTTTTT.TTCCCCCTA.TCTTC.TCTCCT</u>		1	
Haplo_6_74	<u>AATTCGGTTCTTACGTTCTC.TTTTTCTCCCCC.ATTCTCGTCTCCT</u>		1	
			Total 10	

- En base a las frecuencias alélicas, calcular la heterocigosidad H_S de la subpoblación de Pine Forest.
- Calcular el número promedio de diferencias entre pares de alelos π_S en Pine Forest. Para ello, completar primero la siguiente tabla que resume el número de diferencias entre pares de alelos π_{ij} , agregando los valores para las comparaciones entre los haplotipos 1, 2 y 3:

Número de diferencias entre pares de alelos (haplotipos):

	Haplo_1	Haplo_2	Haplo_3	Haplo_4	Haplo_5
Haplo_2		—			
Haplo_3			—		
Haplo_4	40	39	39	—	
Haplo_5	47	46	44	7	—
Haplo_6	48	47	45	12	5

La tabla incluye datos de heterocigosidad incluidos en el artículo, basados en los 11 loci de microsatélites. Completar la tabla con los valores obtenidos en a y b.

	Pine Forest	Warner Mts.
ADN mitocondrial		
H_S		0.64
π_S		22.74
microsatélites		
H_S	0.37	0.78

- c) ¿Cómo pueden interpretarse estos resultados, y qué nos dicen sobre la variación de las subpoblaciones?

2. Estructura dentro de una subpoblación: F_{IS} como descriptor

Consideramos ahora el caso tan simple como importante de la estructura de una subpoblación (o población local) formada por individuos diploides de reproducción sexuada. Nuestros niveles X e Y son, en este caso, los individuos (I) en las subpoblaciones (S).

Sustituyendo los índices en la ecuación [1], obtenemos:

$$t_{IS} = \frac{\bar{t}_S - \bar{t}_I}{\bar{t}_T}$$

Sustituyendo los índices en nuestros estadísticos basados en la variación genética, obtenemos, si usamos clases alélicas (y el modelo de alelos infinitos):

$$F_{IS} = \frac{H_S - H_I}{H_S}$$

[6]

Los ingredientes de esta ecuación deberían resultarnos conocidos. Pensemos en una subpoblación cualquiera $S(1)$:

- $H_{S(1)}$ es la probabilidad de que dos alelos tomados al azar en la subpoblación sean distintos. En el marco de nuestra discusión del modelo de Hardy-Weinberg (HW), le llamamos *heterocigosidad esperada* en la subpoblación $S(1)$. Y $H_{I(1)}$ es la *heterocigosidad observada* en los individuos de la subpoblación $S(1)$.
- Por lo tanto $F_{IS(1)}$ es la diferencia entre la heterocigosidad observada y la esperada en la subpoblación, normalizada al dividir dicha diferencia por la heterocigosidad esperada.
- Si tenemos varias subpoblaciones, F_{IS} será un promedio, reflejando las tendencias medias presentes en las subpoblaciones.

Antes de examinar los factores subyacentes a un cierto valor de F_{IS} para un determinado gen, podemos interpretarlo como un simple descriptor de la variación genotípica dentro de las subpoblaciones:

$F_{IS} > 0$: Necesariamente esto se debe a que $H_S > H_I$: Dentro de las subpoblaciones, hay un déficit de heterocigotas con respecto a lo esperado en base a las correspondientes frecuencias alélicas (dentro de las subpoblaciones).

$F_{IS} = 0$: Dentro de las subpoblaciones, la frecuencia observada de heterocigotas es idéntica a la esperada en base a las frecuencias alélicas. Normalmente, esta es nuestra hipótesis nula: los pares de alelos que forman los genotipos de los individuos son combinaciones al azar de alelos de las subpoblaciones.

$F_{IS} < 0$: Dentro de las subpoblaciones, se observa un exceso de heterocigotas con respecto a lo esperado dadas las frecuencias alélicas locales, i.e., $H_S < H_I$.

La ecuación [4] puede sustituirse por una equivalente usando valores de π :

$$F_{IS} = \frac{\pi_S - \pi_I}{\pi_S}$$

[7]

La interpretación de este descriptor basado en diferencias entre pares de alelos es la misma que hicimos de la ecuación [4].

2.1. Interpretando F_{IS}

Si no hay diferencia entre la heterocigosidad observada y la esperada, $F = 0$, lo esperado bajo el modelo HW. Revisemos brevemente las características de una población de tipo HW:

1- Organismos diploides 2- Reproducción sexual 3- Las frecuencias alélicas no difieren entre sexos 4- Generaciones no solapantes 5- Población de tamaño infinito 6- Apareamientos al azar (los gametos se combinan al azar) 7- No hay migración 8- No hay mutación 9- No actúa la selección natural

Todas las características del 1 al 8 son *demográficas*, ya que tienen que ver con la organización de las poblaciones y su reproducción. Estrictamente, la selección natural es también un proceso demográfico, pero lo separamos de los restantes por su importancia, y también porque los procesos demográficos en sentido estricto afectan a todos los genes que comparten un mismo modo de herencia (por ejemplo, autosómicos, mitocondriales, etc.) del mismo modo. En cambio, la selección es idiosincrática y opera de modos diferentes en cada gen. Solamente en el caso de genes ligados (los del genoma mitocondrial son un caso extremo, ya que no hay recombinación), la selección actuando sobre un gen afecta la variación de otros genes.

Para genes autosómicos, y dejando de lado casos particulares (por ejemplo, zonas de contacto entre grupos de poblaciones muy divergentes, en los que debemos necesariamente considerar la migración), los principales candidatos a generar apartamientos de lo esperado por HW son:

- que los apareamientos no sean al azar;

- que opere la selección natural sobre la variación alélica de nuestro locus de interés.

Obviamente estos dos factores pueden operar sobre un locus de manera simultánea y provocar cambios (con respecto a lo esperado por HW) en la misma dirección, o en direcciones opuestas. Veamos cómo afectan estos factores al coeficiente de endogamia

F_{IS} :

- $F_{IS} = 0$: La reproducción es panmíctica y no hay selección sobre los genotipos. Como es natural, aún en estas condiciones nuestras estimaciones pueden variar ligeramente por azar. Del mismo modo, podemos no ser capaces de detectar apartamientos débiles de la panmixia y selección igualmente débil.
- $F_{IS} < 0$: Los apareamientos son exogámicos (preferentemente unen gametos disímiles más allá de lo esperado por azar) y/o la selección natural favorece a los heterocigotas. Hay un exceso de heterocigotas.
- $F_{IS} > 0$: Los apareamientos son endogámicos y/o la selección natural actúa en contra de los heterocigotas o, lo que es lo mismo, favorece al menos una clase de homocigotas. Hay un déficit de heterocigotas.

Notamos, de paso, que F_{IS} , un coeficiente de endogamia, puede ser positivo por causa de la endogamia, pero también por selección en contra de los heterocigotas, aún en una población panmíctica. De modo análogo, F_{IS} puede ser negativo por causa de la exogamia o por selección a favor de los heterocigotas. La reproducción puede ser panmíctica y producir frecuencias genotípicas en línea con el modelo HW, pero la selección modifica dichas frecuencias, por ejemplo favoreciendo la supervivencia de ciertos genotipos por encima de otros.

Ejemplo 1.

Lemes et al (2014) estudiaron poblaciones humanas brasileñas derivadas de comunidades de esclavos liberados mediante el análisis de SNPs (polimorfismos simples en sitios del genoma) y microsatélites. Estos son valores de F_{IS} para algunos de esos loci:

Locus	F_{IS}
Ace	0.045
Nos3	0.033
Gnb3	-0.007
Agt	-0.041

Salteamos acá cualquier prueba estadística para señalar que, en general, los valores observados se encuentran en un entorno acotado cercano a 0. Puede concluirse que, a nivel local, estos loci no presentan desviaciones significativas de 0. No podemos mostrar que haya desviaciones causadas por apartamientos de la panmixia o la selección natural actuando sobre alguno de estos loci. Este ejemplo muestra valores muy cercanos al de

otras poblaciones humanas, ya que, típicamente, los valores medios de F_{IS} se encuentran en el entorno de ± 0.03 .

Como se ha dicho, la demografía afecta a todos los loci (en este caso autosómicos) por igual, y en este caso puede concluirse que es muy cercana, a nivel local, a la panmixia.

Ejercicio 2: F_{IS} .

La anemia falciforme en humanos es un ejemplo clásico de selección equilibradora (a favor de los heterocigotas) en el locus de la betaglobina. Los heterocigotas AS son resistentes a la malaria; los homocigotas AA son susceptibles a la malaria, y los homocigotas SS padecen anemia falciforme grave. La siguiente tabla resume las frecuencias genotípicas absolutas (número de adultos) observadas en un muestreo a gran escala:

Genotipo	Frecuencia
AA	9365
AS	2993
SS	29
Total	12387

Calcular H_T , H_S y F_{IS} . Explicar qué indica F_{IS} como descriptor de la variación en este locus. ¿Cómo se compara el valor obtenido con el rango usual de F_{IS} en genes humanos? ¿Cómo se explica la diferencia observada?

2.2. Un modelo de endogamia

Hasta aquí presentamos al estadístico F_{IS} , en primer lugar como un *descriptor* de la estructura de una población local, independientemente de los factores subyacentes a los valores observados. En segundo lugar, discutimos los principales factores que pueden resultar en un F_{IS} distinto de 0. Para un locus aislado, tanto la selección natural como el régimen de apareamientos (y más en general las características demográficas de las poblaciones), solos o en combinación, pueden producir valores distintos de 0. Para muchos loci, podemos detectar un patrón general atribuible a la demografía. En humanos, F_{IS} es muy cercano a 0, sugiriendo que las poblaciones locales están muy cerca de un régimen de apareamientos al azar (con respecto a la variación genética). Por contraste, un gen como el de la beta globina en regiones históricamente afectadas por la malaria muestra F_{IS} que se apartan significativamente de esa norma, reflejando la acción y el tipo de selección natural sobre este locus en particular.

En esta sección, veremos un modelo sencillo de población con endogamia. Consideramos un locus autosómico diploide, con dos alelos (A y a) neutrales (los 3 genotipos son equivalentes en términos de su eficacia darwiniana). La única diferencia entre este modelo y el de HW es que los gametos no se combinan de manera enteramente aleatoria para formar los organismos diploides, sino que existe cierta tendencia a que los alelos del mismo tipo se combinen en una frecuencia mayor a la esperada por azar. Para modelar esta circunstancia, definimos la siguiente situación:

- una fracción F de los gametos se aparean entre homónimos
- una fracción $(1-F)$ se aparean de manera aleatoria

Notamos que si $F = 0$, la población es de tipo HW. ¿Cuáles son las frecuencias genotípicas esperadas en cada caso?

Genotipo	Modelo HW	Modelo con endogamia
AA	p^2	$p^2(1 - F) + pF$
Aa	$2pq$	$2pq(1 - F)$
aa	q^2	$q^2(1 - F) + qF$

Para estimar F , observamos que $H_I = 2pq(1 - F)$ y que $2pq = H_S$

En palabras, la heterocigosidad observada en los individuos de la subpoblación puede ser utilizada para estimar F . Notemos que la heterocigosidad observada está reducida en una fracción F con respecto al valor esperado bajo HW. Reemplazando $2pq$ por H_S y despejando F , obtenemos:

$$F = \frac{H_S - H_I}{H_S}$$

F de este modelo es nuestro F_{IS} . Comprobamos que, para este modelo, F_{IS} es la fracción de gametos que se aparean con gametos del mismo tipo. Históricamente, F es el coeficiente de endogamia original para una única población. La generalización a diversos estadísticos F en poblaciones estructuradas llevó a la necesidad de introducir subíndices para especificar qué niveles de organización se están considerando.

En general, si estamos considerando genes neutrales (o al menos cuasi neutrales, que asimilamos a la neutralidad para nuestros propósitos) y observamos un valor de $F = F_{IS}$ de, por ejemplo, 0,10, podemos decir que el patrón reproductivo de la población, seguramente complejo, resulta en un déficit de heterocigotas del 10% con respecto a HW. En este sentido, la heterocigosidad observada es equivalente a la de un modelo con endogamia donde una fracción $F = 0,10$ de los alelos se aparean con los de su clase, mientras que el 90% restante se aparean al azar y resulta en proporciones HW.

2.3. Comentarios adicionales sobre F_{IS}

En general, la identificación de unidades discretas dentro de las poblaciones es relativamente incierta. Las poblaciones pueden organizarse de manera continua en el espacio, o el investigador define unidades de manera más o menos arbitraria (colectas realizadas desde una misma base de campo, por ejemplo). Los *organismos* o *individuos* son, en cambio, unidades naturales, y para sistemas diploides los pares de alelos de los individuos constituyen sus genotipos. Por tanto, $H_I = H_o$, la frecuencia observada de heterocigotas. Este es el único nivel en el que la heterocigosidad se observa de manera directa en los individuos. Cuando X es, por ejemplo, el nivel de las subpoblaciones S , H_S , la probabilidad de que dos alelos tomados al azar dentro de las subpoblaciones sean distintos, es una propiedad de la subpoblación, no de los individuos que la forman. Otra forma de plantearlo es que F_{XY} involucra heterocigosidades observadas solamente

cuando el nivel X es el de los individuos. Esto incluye claramente a F_{IS} , pero no a F_{ST} , que ocupará buena parte de la discusión que sigue.

Una observación vinculada es que los estadísticos F se aplican a organismos haploides y genes de transmisión uniparental (mitocondriales, del cromosoma Y), siempre y cuando no se considere el nivel I , que solamente tiene sentido cuando hay genotipos diploides.

3. Variación geográfica

Todas las consideraciones de la Sección 1 (Estructura poblacional) se aplican a la variación geográfica. Su estudio admite diversos niveles de organización, pero vamos a considerar primero la más sencilla: la de subpoblaciones anidadas en la población total. En este caso, el nivel inferior $X = S$ y el superior $Y = T$. Podría haber, en este caso, k subpoblaciones tales que $S_1, S_2, \dots, S_k \in T$. Al igual que en la sección previa, aspiramos a obtener un estimador de la diferencia entre la edad promedio de pares de alelos del nivel S .

El estimador genealógico de estructura poblacional de la ecuación [1] toma, en este contexto, la siguiente forma:

$$t_{ST} = \frac{\bar{t}_T - \bar{t}_S}{\bar{t}_T}$$

[8]

Si queremos utilizar medidas de heterocigosidad (y el modelo de alelos infinitos), nuestro estimador será:

$$F_{ST} = \frac{H_T - H_S}{H_T}$$

[9]

Y, claro está, para usar diferencias entre las secuencias de pares de alelos recurriremos a:

$$F_{ST} = \frac{\pi_T - \pi_S}{\pi_T}$$

[10]

Por las dudas, recordemos que π , el número promedio de diferencias entre pares de alelos está definido en la ecuación [4]:

$$\hat{\pi} = \sum_{i < j}^k 2p_i p_j \hat{\pi}_{ij}$$

[4]

Usaremos esta ecuación separadamente para los valores de cada subpoblación, así como para la población total. En cada caso, naturalmente, utilizamos las clases alélicas correspondientes y sus frecuencias.

F_{ST} , conocido como *índice de fijación*, es uno de los estadísticos más utilizados en genética de poblaciones.

3.1. F_{ST} como descriptor de la estructura geográfica

Aunque la estructura geográfica puede tomar varias formas, una característica general es que los individuos dentro de las subpoblaciones tienen más probabilidad de aparearse entre sí que individuos tomados al azar en toda la distribución. Como consecuencia, los pares de alelos tomados al azar dentro de las subpoblaciones estarán, en promedio, más emparentados que pares de alelos tomados al azar de la población total (o, lo que es casi lo mismo si hay muchas subpoblaciones, tomados al azar de distintas subpoblaciones). Es decir que $t_T \geq t_S$.

Esto implica que $F_{ST} \geq 0$. Esta condición, combinada con el hecho de que F_{ST} está normalizado, lleva a que, necesariamente, $0 \leq F_{ST} \leq 1$. Si $F_{ST} = 0$, no hay diferencias entre las subpoblaciones y la población total. Las subpoblaciones son iguales entre sí del punto de vista genético y se presentan como réplicas de la población total. Tomar alelos al azar de la población total es equivalente a tomarlos dentro de las subpoblaciones.

En cambio, si $F_{ST} > 0$, será un indicador del grado de diferenciación, en promedio, de las subpoblaciones.

Ejemplo 2

Mora et al (2007) compararon la estructura poblacional de dos especies de tucu-tucus a lo largo de la costa de la Provincia de Buenos Aires: *Ctenomys australis* está vinculado a la primera línea de dunas costeras, y es una especie de mayor tamaño y capacidad de dispersión; *C. talarum* ocupa suelos arenosos paralelos a las dunas (aunque también hay poblaciones en áreas alejadas de la costa). En cada localidad de muestreo, se tomaron muestras de las dos especies, cubriendo una distancia de unos 300 km. Los resultados fueron los siguientes:

Especie	F_{ST}
<i>C. australis</i>	0,27
<i>C. talarum</i>	0,79

Las dos especies tienen una subdivisión geográfica significativa, como es esperable para roedores subterráneos con capacidad de dispersión limitada. Sin embargo, observamos que la especie con menor capacidad de dispersión tiene una subdivisión geográfica mucho más marcada. Veremos luego que es posible que el hábitat ocupado por *C. australis* ha sido, históricamente, más inestable, y la especie ha debido (re)colonizar la región en tiempos recientes, posiblemente durante el Holoceno.

Versatilidad del F_{ST}

El uso de este estadístico requiere solamente definir dos niveles anidados de estructura poblacional S y T . En la **Fig. 2** se muestran gráficamente algunos ejemplos:

- F_{ST} clásico: se considera un conjunto de subpoblaciones dentro de una población total;

- F_{ST} entre pares de poblaciones: para cada comparación, se consideran solamente dos subpoblaciones, y la población total simplemente las combina en un todo;
- F_{ST} propio de cada población: se consideran las poblaciones una a una, y en cada caso se las compara con la población total.

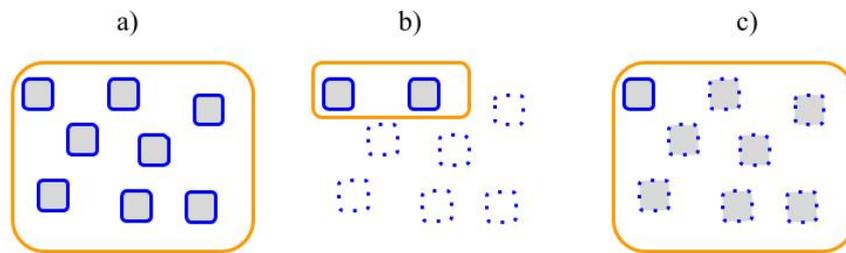


Fig 2. Distintos usos del índice de fijación F_{ST} a) índice clásico, considerando un conjunto de subpoblaciones y su divergencia promedio con respecto a la población total que las comprende; b) F_{ST} entre pares de poblaciones; para cada par (la figura muestra un par como ejemplo), se consideran dos subpoblaciones, y la población total formada por ambas; c) F_{ST} propio de cada subpoblación, considerando su divergencia con respecto a la población total, que en este caso es igual a la usada en a).

Ejemplo 3

Bell et al. (2021) estudiaron la variación geográfica en una especie de ardilla (*Tamias amoenus*) en el suroeste de Estados Unidos, genotipando 11 loci de microsatélites en varias subpoblaciones. El objetivo principal era estudiar el nivel de divergencia entre subpoblaciones y saber si una de ellas (Pine Forest) era más divergente que las demás, teniendo en cuenta su aislamiento geográfico.

Por otra parte, se plantearon establecer si la variación dentro de esa subpoblación era comparable a la de otras estudiadas o estaba reducida. En el *Ejercicio 1* ya comparamos dos medidas de variación local (H_S y π_S) en Pine Forest y Warner Mts.

F_{ST} entre pares de poblaciones de *Tamias amoenus*, en base a loci de microsatélites.

	Pine Forest	Warner Mts.	Galena	Boca Rsvr.	Twin Lakes
Pine Forest	—				
Warner Mts.	0.525	—			
Galena	0.354	0.091	—		
Boca Rsvr.	0.304	0.096	0.012	—	
Twin Lakes	0.382	0.127	0.047	0.045	—
\bar{H}_S	0.37	0.78	0.75	0.75	0.73

Observamos que \bar{H}_S , promediada para los 11 loci (recordemos, de paso, que \bar{H}_S es la heterocigosidad esperada por HW dentro de cada subpoblación), en Pine Forest es aproximadamente la mitad de los valores de las restantes subpoblaciones.

En cuanto a la estructura geográfica, los F_{ST} entre pares de subpoblaciones varían de valores bajos (0,045) a moderados (0,127) entre subpoblaciones, a excepción de las comparaciones que involucran a Pine Forest, que son notoriamente más altos (0,304-0,525). En suma, esta subpoblación es más divergente y tiene menor variación genética que las restantes.

3.2. Interpretando F_{ST} : primera aproximación

Al igual que en el caso de F_{IS} , para F_{ST} la demografía afecta a todos los loci, mientras que la selección natural es idiosincrática. Por lo tanto, interpretamos las tendencias que son consistentes para múltiples loci (en especial si no son candidatos a estar bajo selección fuerte) como reflejo de la demografía. Un locus que se aparta fuertemente de la línea de base demográfica es sospechoso de estar bajo selección natural positiva, y ésta puede favorecer o limitar la divergencia entre poblaciones.

La gran diferencia entre F_{IS} y F_{ST} es que el primero compara la ancestría de alelos de los individuos (los alelos de los genotipos observados) con la de alelos al azar en las subpoblaciones. Sabemos que los genotipos de una generación son el resultado de los apareamientos de la generación precedente. Cualquiera sea el régimen de apareamientos, se reflejará en forma “instantánea” en los genotipos observados. La demografía que interesa para interpretar F_{IS} es la de la generación precedente a la actual. Y lo mismo se aplica para la selección natural: solamente afectará F_{IS} si está operando en la actualidad para distorsionar las frecuencias genotípicas.

En cambio, F_{ST} depende de la historia demográfica de las poblaciones, que involucra, para genes neutrales, la acción de la divergencia por deriva genética entre poblaciones y el papel del flujo génico como moderador de dicha divergencia. Esos procesos demográficos (deriva genética y flujo génico) operan a ritmos muy lentos, de modo que lo que observamos hoy se verá influenciado por la historia de las poblaciones.

Dos poblaciones pueden ser poco divergentes (F_{ST} cercano a 0) porque intercambian muchos migrantes (flujo génico alto), pero también porque se han dividido hace muy poco tiempo, aún si el flujo génico ha sido bajo o nulo desde la división de las poblaciones.

Antes de estudiar el proceso de divergencia y el papel del flujo génico, vale la pena hacer un comentario adicional para resaltar que F_{IS} y F_{ST} son completamente independientes. Las subpoblaciones pueden ser completamente panmícticas a nivel local ($F_{IS} = 0$) y estar muy diferenciadas geográficamente (F_{ST} alto), por ejemplo. Cualquier combinación de valores de los dos estadísticos es en principio posible, y se interpretan en forma independiente.

Ejercicio 3: F_{ST}

Retomamos el ejemplo de las ardillas (*Tamias amoenus*) estudiadas por Bell et al. (2021). El artículo reporta que el F_{ST} promedio de 11 loci de microsatélites entre Pine Forest y Warner Mts. es el mayor del estudio, $\bar{F}_{ST} = 0.525$. Este es un valor que corresponde a subpoblaciones muy diferenciadas geográficamente.

Utilizando los datos mitocondriales presentados en el Ejercicio 2, obtenemos los siguientes valores de divergencia media entre pares de alelos dentro de las subpoblaciones y en el total (que, en este caso, combina las dos subpoblaciones de interés; asumiendo que las subpoblaciones son iguales, las frecuencias de los haplotipos en la población total serán el promedio de las frecuencias en las subpoblaciones):

Pine Forest:

$\pi_S = 0,82$ (obtuvimos este resultado en el Ejercicio 2) Warner Mts.:

$\pi_S = 22,74$

Total: $\pi_T = 16,70$

Estimar F_{ST} (que, en este caso, podríamos etiquetar como π_{ST} o $F_{ST(\pi)}$ si queremos destacar que la estimación se basa en el número de diferencias pareadas entre alelos) en base a estos valores. ¿Cómo se interpreta, y cómo se compara con el valor de F_{ST} reportado en el artículo en base a microsatélites ($F_{ST} = 0,525$; según tabla del Ejemplo 3 más arriba).

3.3. Equilibrio entre deriva local y migración en modelos de islas

Vamos a considerar un modelo sencillo de estructura poblacional: la población total está dividida en d subpoblaciones (islas), cada una de las cuales tiene N individuos diploides. En cada generación, una fracción m de los individuos de cada isla es reemplazada por individuos migrantes de cualquiera de las demás poblaciones. Observamos que todas las islas son iguales, no solamente en el tamaño de las subpoblaciones que componen, sino en cuanto al régimen de intercambio de migrantes. El fenómeno de la migración se llama también flujo génico, para enfatizar que los individuos se incorporan reproductivamente a las poblaciones de recepción, a las que aportan, por tanto, sus genes.

Modelo de islas

Quedan para más adelante la demostración de que, en un modelo de islas, con un gran número de islas (en el límite, infinitas):

$$F_{ST} \approx \frac{1}{4Nm + 1} \quad [10]$$

Por el momento, observamos lo siguiente:

- un modelo de islas supone que la población total está dividida en subpoblaciones, todas ellas de igual tamaño poblacional local N
- el régimen de migración es el mismo para todas las islas: cada una recambia una fracción m de sus N individuos en cada generación;
- m es la tasa de migración o flujo génico, y cada migrante tiene igual probabilidad de provenir de cualquiera de las otras islas

Para genes neutrales, la divergencia entre subpoblaciones en este modelo depende de dos procesos contrapuestos:

- la deriva genética local, que ocurre a un ritmo $1/2N$ en cada generación; para dos alelos tomados de una misma subpoblación;
- la migración o flujo génico, cuya intensidad está reflejada en m

Observamos cómo, en un modelo de islas, F_{ST} captura el balance entre esos dos procesos contrapuestos en la ecuación [10], sin que ello la demuestre:

Tomamos dos alelos al azar en una subpoblación y seguimos su ancestría hacia el pasado.

- En cada generación, la probabilidad de que los dos alelos provengan de un mismo alelo ancestral (probabilidad de coalescencia) es $1/2N$.
- Por otra parte, la probabilidad de que uno cualquiera de los dos alelos provenga de un migrante en la generación anterior es m . Tomados en conjunto, la probabilidad de migración para el par es $2m$. Despreciamos la probabilidad de que los dos migren en una misma generación ($2m^2$).
- también despreciamos la probabilidad de que, en una misma generación, ocurra una coalescencia y una migración ($m/2N$). Estas aproximaciones son válidas si m es pequeño y $2N$ es grande. En suma, tomamos los dos eventos (coalescencia y migración) como mutuamente excluyentes (Fig. 3)

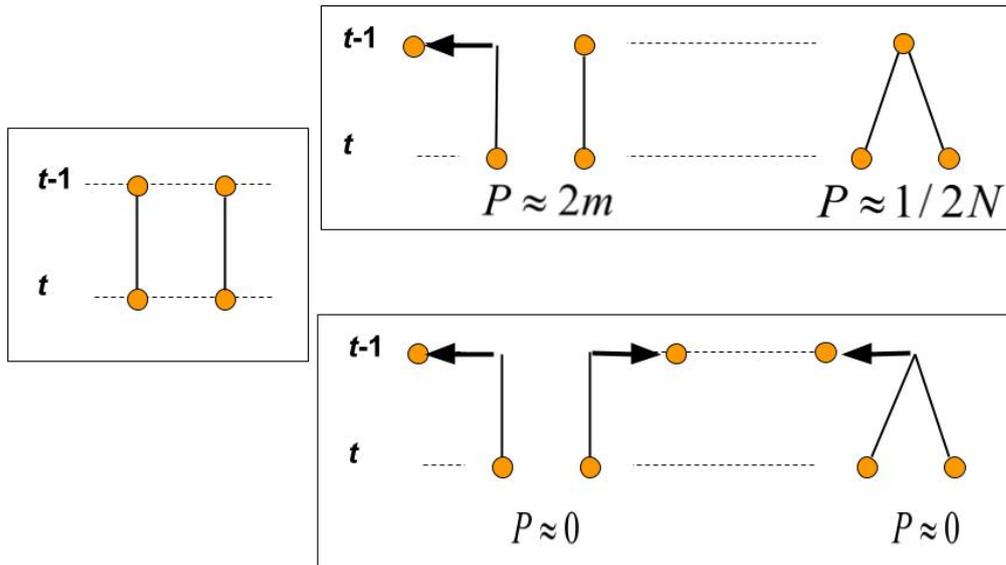


Fig. 3. Esquema ilustrando las probabilidades de migración y coalescencia (en ambos casos, en una generación hacia el pasado) para un par de alelos tomados dentro de una subpoblación (panel superior). Despreciamos la probabilidad de que, en una generación, ocurran dos eventos (panel inferior). Si en una generación no ocurre ni una coalescencia ni un evento de migración (panel izquierdo), nos movemos a t-1 y las probabilidades de los eventos relevantes (panel superior) se mantienen.

El balance entre deriva local y migración determina cuánto menor será \bar{t}_S que \bar{t}_T , es decir el balance entre $1/2N$ y $2m$:

$$F_{ST} \approx \frac{1/2N}{2m + 1/2N}$$

Multiplicando numerador y denominador por $2N$, obtenemos la ecuación [10].

Observamos que este balance entre deriva local y migración es análogo al balance entre deriva y mutación, que obtuvimos con un razonamiento semejante:

$$F_{DM} \approx \frac{1}{4N\mu + 1}$$

F_{DM} y su complemento H_{DM} son valores esperados en una población como resultado del equilibrio dinámico entre deriva genética y mutación (neutral) para un modelo de Wright-Fisher con una población de tamaño constante N . Sabemos que, si por alguna razón la población no está en equilibrio (por ejemplo, ocurrió un cambio de tamaño poblacional), el arribo al nuevo equilibrio llevará mucho tiempo (del orden de $2N$ generaciones), porque depende de dos procesos (deriva y mutación) que ocurren a tasas muy bajas ($1/2N$ y μ , respectivamente). El valor de F_{DM} en un momento dado reflejará el comportamiento de la población a lo largo de un período largo de tiempo.

Del mismo modo, F_{ST} es un valor esperado, en equilibrio entre deriva y migración, en un modelo de islas. Nuevamente, si la población no está en equilibrio (por cambios en su estructura o dinámica demográfica), el arribo al equilibrio será típicamente lento,

porque depende de la deriva local y la migración, con el único matiz de que m puede ser mayor que μ .

Ejercicios resueltos

Ejercicio 1: variación en subpoblaciones

- a) En base a las frecuencias alélicas, calcular la heterocigosidad H_S de la subpoblación de Pine Forest.

Usamos las frecuencias de los tres alelos presentes en la subpoblación de Pine Forest: $f(\text{haplo}_1)=0.1$, $f(\text{haplo}_2)=0.7$, y $f(\text{haplo}_3)=0.2$. La heterocigosidad de la subpoblación

es $\hat{H} = \sum_{i < j}^k 2p_i p_j$. En este caso $H_S = 0,46$

- b) Calcular el número promedio de diferencias entre pares de alelos π_S en Pine Forest. Para ello, completar primero la siguiente tabla que resume el número de diferencias entre pares de alelos π_{ij} , agregando los valores para las comparaciones entre los haplotipos 1, 2 y 3:

Observando las secuencias de Pine Forest, las diferencias observadas entre los haplotipos son: haplo_1 vs. haplo_2: 1 haplo_1 vs. haplo_3: 3 haplo_2 vs. haplo_3: 2

Completando la tabla correspondiente obtenemos:

Número de diferencias entre pares de alelos (haplotipos):

	Haplo_1	Haplo_2	Haplo_3	Haplo_4	Haplo_5
Haplo_2	1	---			
Haplo_3	3	2	---		
Haplo_4	40	39	39	---	
Haplo_5	47	46	44	7	---
Haplo_6	48	47	45	12	5

Con estos valores, podemos completar la tabla de diferencias entre haplotipos. Para calcular π_S en Pine Forest, usamos la ecuación [4]. Observamos que en la parte a sumamos los " $2p_i p_j$ ", o sea los productos de las frecuencias de los alelos (haplotipos). Ahora, a cada término " $2p_i p_j$ " lo multiplicamos por el número de diferencias entre los alelos i y j . Usamos las diferencias de las secuencias de los alelos, y no solamente sus frecuencias, para este estimador de variación genética.

Obtenemos $\pi_S = 0,82$ para Pine Forest. Vale comentar que esto significa que dos pares de alelos en esta subpoblación tienen, en promedio, algo menos que 1 diferencia en este locus. Ese promedio incluye pares que no tiene diferencias (por ejemplo, por azar tomo dos alelos de una misma clase, digamos haplo_1), otras que tienen 1, diferencia, etc.

La tabla de estadísticos de variación poblacional, una vez completada, es la siguiente:

	Pine Forest	Warner Mts.
ADN mitocondrial		
H_S	0.46	0.64
π_S	0.82	22.74
microsatélites		
H_S	0.37	0.78

- c) ¿Cómo pueden interpretarse estos resultados, y qué nos dicen sobre la variación de las subpoblaciones?

La principal observación es que los estimadores de variación mitocondrial obtenidos son consistentes con los valores de H_S reportados en el artículo en base a 11 loci de microsatélites: también en el ADN mitocondrial encontramos que la variación de la subpoblación de Pine Forest es mucho menor que la de Warner Mts.

Nuestras estimaciones basadas simplemente en clases alélicas (H_S) no son muy confiables porque no tienen en cuenta la divergencia entre los alelos. Si miramos las secuencias presentadas en el ejercicio o la tabla de diferencias entre pares de alelos, algunos haplotipos tienen 1 o unas pocas diferencias, mientras que otros difieren en casi todos los sitios variables.

Aquellos basados en π_S hacen mejor uso de los datos. En ellos, observamos más de 20 veces más variación en Warner Mts. que en Pine Forest. Recordemos que las secuencias utilizadas tenían 620 pares de bases. Pasando los valores de π_S a %, obtenemos:

$$\pi_S \text{ Pine Forest} = 0.11\% \quad \pi_S \text{ Warner Mts} = 3.16\%$$

En promedio, dos secuencias tomadas al azar en Pine Forest tienen menos de 0.2% de divergencia, mientras que en Warner Mts. tienen más de 3% de divergencia.

Ejercicio 2: F_{IS}

Primero, estimamos $p = f(A) = 0,877$, y $q = 0,123$. La frecuencia esperada de heterocigotas (HW) es $\hat{H}_S = \hat{H}_e = 0,2157$, mientras que la frecuencia observada es $H_I = H_o = 0,2416$. En consecuencia, $F_{IS} = -0,120$. En humanos, los valores típicos de $F_{[IS]}$ son cercanos a 0, de modo que este locus se aparta significativamente, y observamos un 12% de exceso de heterocigotas con respecto a HW. Descartamos que ese 12% se deba a exogamia, ya que la exogamia se reflejaría en todos los loci autosómicos de igual manera. Concluimos que el exceso de heterocigotas se debe a selección natural en favor de los mismos. En este caso, sabemos que los heterocigotas son resistentes a la malaria y, a diferencia de los homocigotas SS , no presentan anemia falciforme.

Ejercicio 3: F_{ST} .

Retomamos el ejemplo de las ardillas (*Tamias amoenus*) estudiadas por Bell et al. (2021). El artículo reporta que el F_{ST} promedio de 11 loci de microsatélites entre Pine Forest y Warner Mts. es el mayor del estudio, $F_{ST} = 0,525$. Este es un valor que corresponde a subpoblaciones muy diferenciadas geográficamente.

Utilizando los datos mitocondriales presentados en el Ejercicio 2, obtenemos los siguientes valores de divergencia media entre pares de alelos dentro de las subpoblaciones y en el total (que, en este caso, combina las dos subpoblaciones de interés; asumiendo que las subpoblaciones son iguales, las frecuencias de los haplotipos en la población total serán el promedio de las frecuencias en las subpoblaciones):

Pine Forest:

$\pi_S = 0,82$ (obtuvimos este resultado en el Ejercicio 2) Warner Mts.:

$\pi_S = 22,74$

Total: $\pi_T = 16,70$

Estimar F_{ST} (en este caso igual a π_{ST} si queremos destacar que la estimación se basa en el número de diferencias pareadas entre alelos) en base a estos valores. ¿Cómo se interpreta, y cómo se compara con el valor de F_{ST} reportado en el artículo en base a microsatélites ($F_{ST} = 0,525$; según tabla del Ejemplo 3 más arriba).

Para estimar F_{ST} , comenzamos por calcular el valor medio de $\pi_S = 11,78$. Utilizando la ecuación [10], obtenemos $F_{ST} = 0,295$. Esto indica que, en promedio, la divergencia entre pares de alelos tomados dentro de las subpoblaciones es 30% menor que la divergencia entre alelos tomados de la población total. Recordemos que este es un F_{ST} entre pares de poblaciones. Esta estimación está en línea con la idea de que estas dos subpoblaciones están bien diferenciadas genéticamente.

La estimación mitocondrial es menor que la obtenida con microsatélites. Esa diferencia puede deberse a que en nuestro ejercicio usamos un subconjunto de los datos (mayor error de estimación), o puede deberse a diferencias reales entre la divergencia mitocondrial y la nuclear. La basada en microsatélites tiene la ventaja de ser un promedio de 11 loci, mientras que, al no haber recombinación, todo el genoma mitocondrial se comporta como 1 solo locus. Sería interesante saber si los π_{ST} pareados para las restantes poblaciones siguen un patrón comparable a lo reportado para microsatélites.

Referencias

Bell, K. C., J. Van Gunst, M. B. Teglás, J. Hsueh, M. D. Matocq. 2021. Lost in a sagebrush sea: comparative genetic assessment of an isolated montane population of *Tamias amoenus*. Journal of Mammalogy, 102:173–187. [DOI:10.1093/jmammal/gyaa166](https://doi.org/10.1093/jmammal/gyaa166)

Kingman, J. F. C. 1982. The coalescent. Stochastic Processes and Their Applications 13:235-248. [https://doi.org/10.1016/0304-4149\(82\)90011-4](https://doi.org/10.1016/0304-4149(82)90011-4)

Lemes, R. B., K. Nunes, D. Meyer, R. C. Mingroni-Netto, Paulo A. Otto. 2014. Estimation of Inbreeding and Substructure Levels in African-Derived Brazilian Quilombo Populations. Human Biology 86:276-288. doi: 10.13110/humanbiology.86.4.0276.

Slatkin, M. 1991. Inbreeding coefficients and coalescence times. *Genetics Research*, Cambridge 58: 167-175. [doi:10.1017/s0016672300029827](https://doi.org/10.1017/s0016672300029827).

Wright, S. 1951. The genetical structure of populations. *Annals of Eugenics* 15: 323-354.