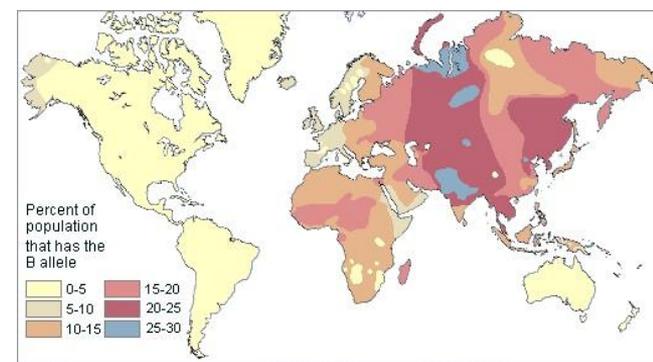


Curso de Evolución 2024
Facultad de Ciencias
Montevideo, Uruguay

<http://eva.fcien.universidad.edu.uy/>

<https://www.youtube.com/@CursoEvolucion/videos>



7. Variación geográfica. Estructura poblacional y flujo génico. Estadísticos F y su interpretación. Equilibrio entre deriva y flujo génico.

Nuestros dos modelos de referencia (HW, WF)

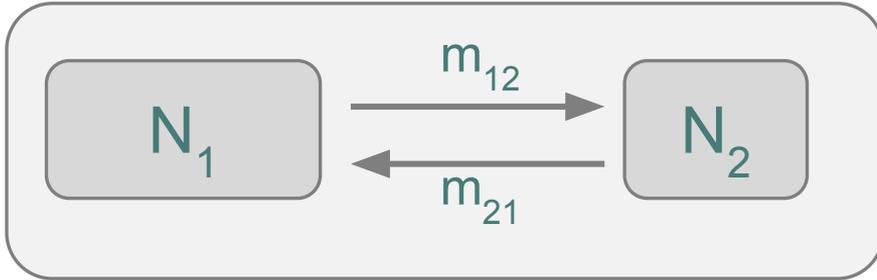
- Más allá de sus diferencias, las poblaciones son “panmícticas” (homogéneas):
 - en HW eso significa que los apareamientos ocurren al azar y que los individuos (diploides) son combinaciones aleatorias de alelos neutrales.
 - En WF eso se manifiesta que todos los pares de alelos tienen la misma probabilidad de ser descendientes de un mismo ancestro.
- En ambos casos, los alelos son intercambiables por su equivalencia demográfica (parentesco, combinaciones genotípicas) y funcional (neutralidad).
- En su forma básica, estos modelos no admiten inmigración. Las poblaciones son internamente homogéneas e independientes unas de otras.
- Al introducir estructura poblacional, perdemos la intercambiabilidad de los alelos (o al menos la condicionamos a su emplazamiento) e introducimos dos conexiones **potenciales** entre poblaciones (o subpoblaciones):
 - Conexión ancestral: las poblaciones de interés tienen una población ancestral en común.
 - Conexión actual (o más en general posterior a la separación del ancestro): migración o flujo génico.
- Hasta ahora, consideramos procesos en una dimensión (tiempo) con un orden natural; al introducir estructuras espaciales, sociales, etc., agregamos dimensiones con un comportamiento menos evidente.

Poblaciones estructuradas: ideas básicas

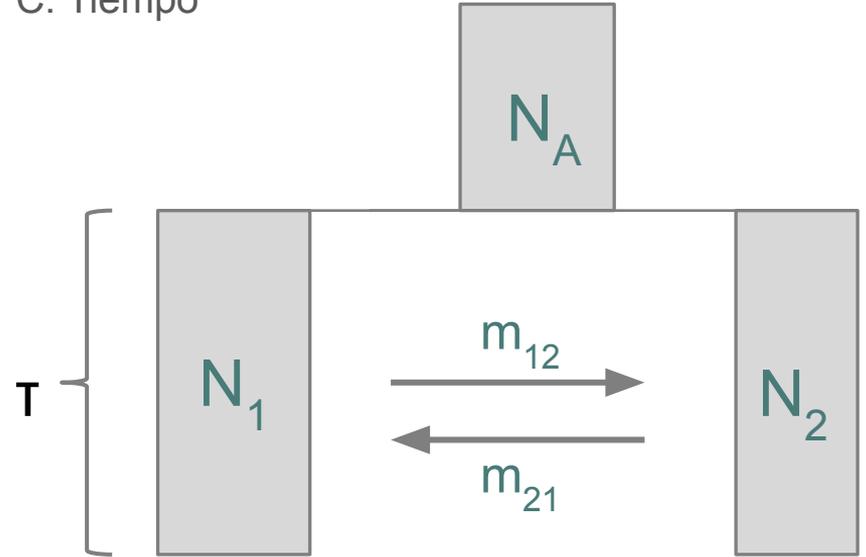
A. Estructura



B. Espacio

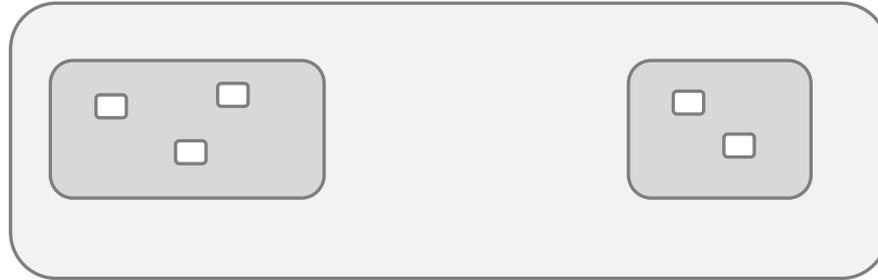


C. Tiempo



Estructura poblacional: descripción genealógica

A. Estructura



- Comenzamos por un enfoque descriptivo (en términos genéticos) y genealógico (parentesco entre individuos o alelos).
- Nota: no seguimos el curso de la historia del tema, sino que examinamos el problema con las ventajas del enfoque genealógico.

Estructura poblacional: descripción genealógica

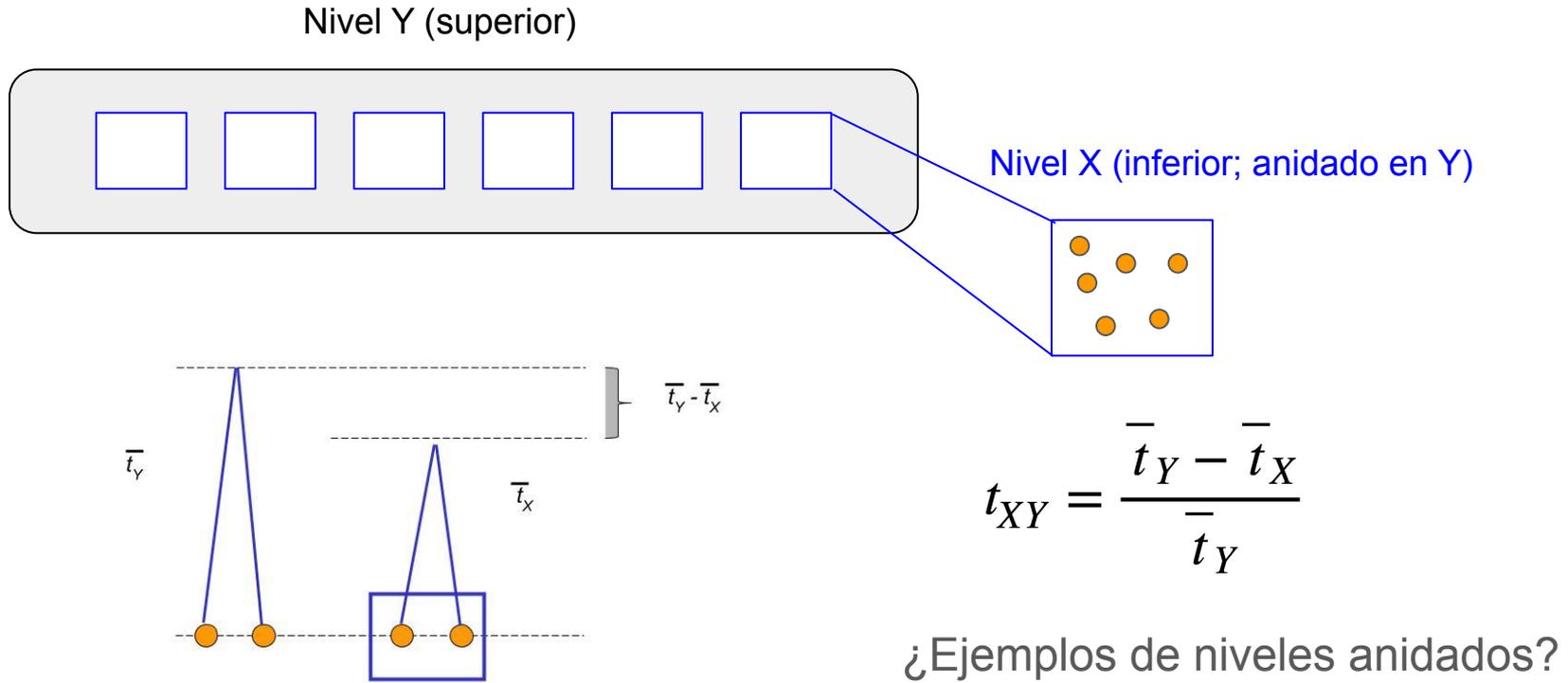


Fig 1. Antigüedad media del ancestro común de dos pares de alelos tomados en un nivel superior (Y) y en un nivel inferior (X) anidado en el anterior.

Estructura poblacional: descripción genealógica

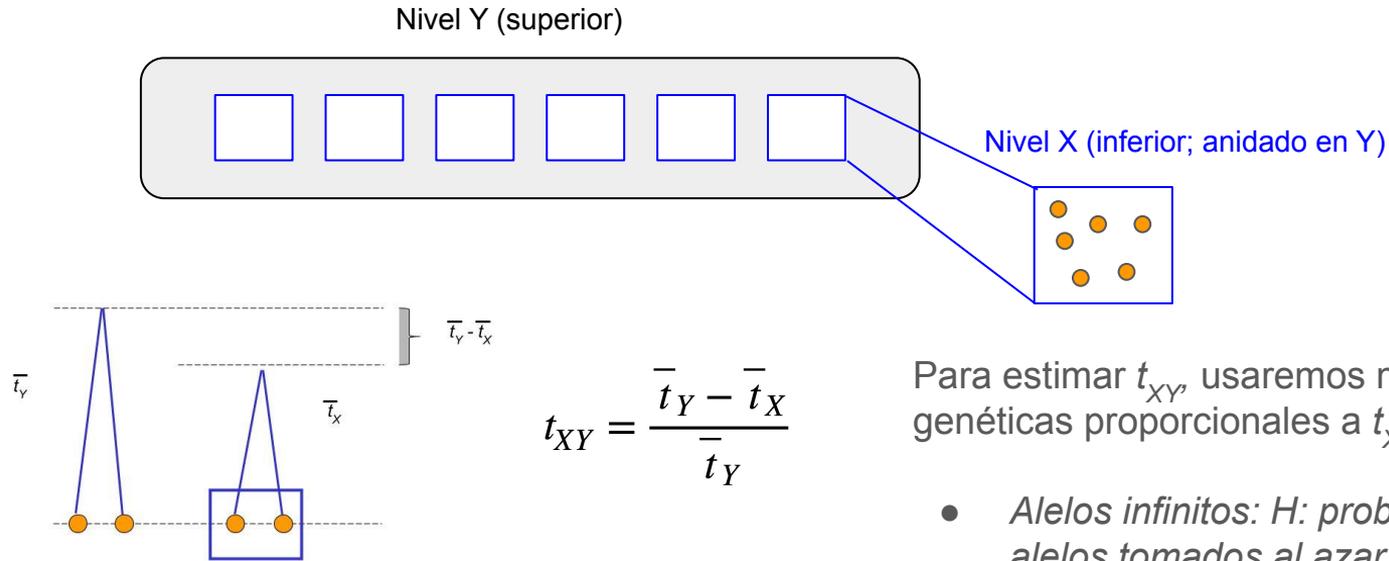


Fig 1. Antigüedad media del ancestro común de dos pares de alelos tomados en un nivel superior (Y) y en un nivel inferior (X) anidado en el anterior.

Para estimar t_{XY} , usaremos medidas genéticas proporcionales a t_X y t_Y :

- *Alelos infinitos*: H : prob. de que 2 alelos tomados al azar sean diferentes
- *Sitios infinitos*: π : número de diferencias entre pares de alelos

Estructura poblacional: descripción genealógica

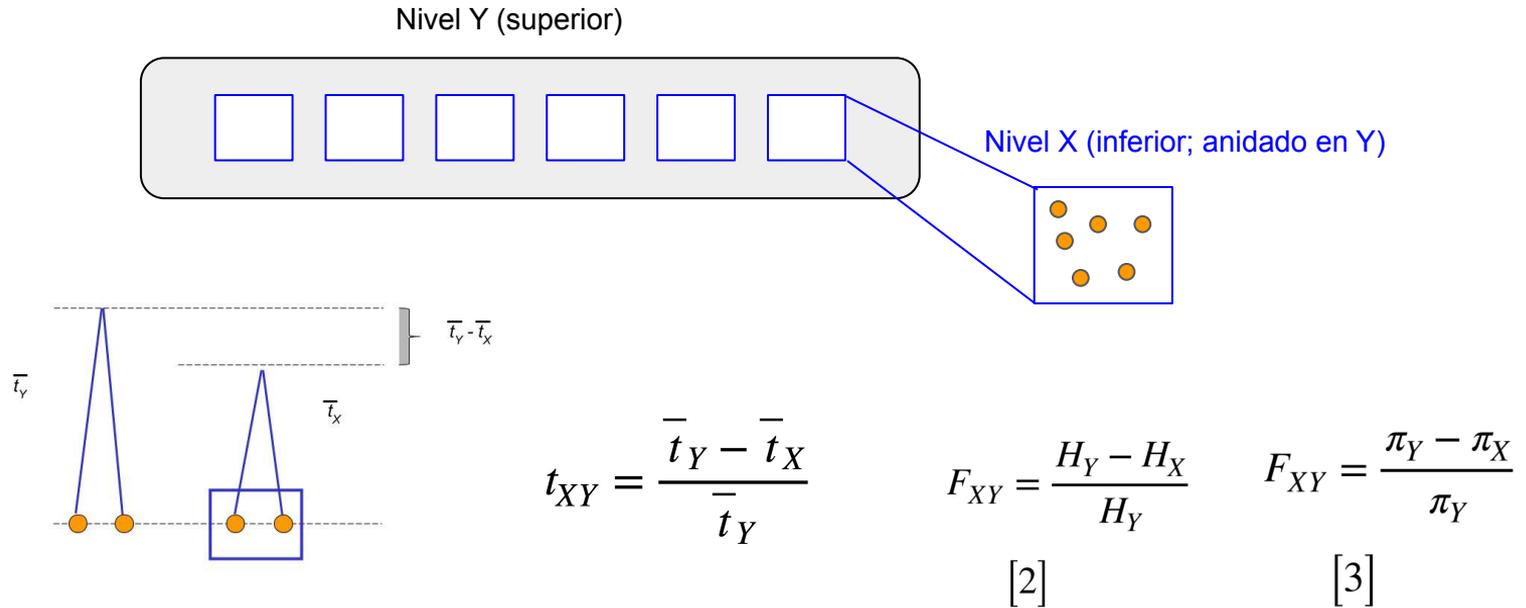


Fig 1. Antigüedad media del ancestro común de dos pares de alelos tomados en un nivel superior (Y) y en un nivel inferior (X) anidado en el anterior.

Ejercicio 1: variación en subpoblaciones de ardillas (*Tamias amoenus*)



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 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%.

Ver material de apoyo: 5. Coeficientes de endogamia y estructura poblacional



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%.

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

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

	Pine forest	Warner Mts.	Total	
Haplo_1_00	GGCCTTACCTCCTACCTCTCCCCCTCCTTTTTTCGCATCCTACTCTTC	1	5	6
Haplo_2_94C.....	7		7
Haplo_3_06C.....A.....T.....	2		2
Haplo_4_14	AAT.CCGTTCTTAC..TC.CTTTTT.TTCCC.CCTA.TCTTC..CT.CT	3		3
Haplo_5_73	AATTCCGTTCTTACGTTCTCTTTTT.TTCCCCCTA.TCTTC.TCTCCT	1		1
Haplo_6_74	AATTCCGTTCTTACGTTCTC.TTTTTCTCCCCC.ATTCTTCGTCTCCT	1		1
	Total	10	10	20
Pine forest				
Haplo_1_00	GGCCTTACCTCCTACCTCTCCCCCTCCTTTTTTCGCATCCTACTCTTC		1	
Haplo_2_94C.....		7	
Haplo_3_06C.....A.....T.....		2	
	Total		10	
Warner Mts.				
Haplo_1_00	GGCCTTACCTCCTACCTCTCCCCCTCCTTTTTTCGCATCCTACTCTTC		5	
Haplo_4_14	AAT.CCGTTCTTAC..TC.CTTTTT.TTCCC.CCTA.TCTTC..CT.CT		3	
Haplo_5_73	AATTCCGTTCTTACGTTCTCTTTTT.TTCCCCCTA.TCTTC.TCTCCT		1	
Haplo_6_74	AATTCCGTTCTTACGTTCTC.TTTTTCTCCCCC.ATTCTTCGTCTCCT		1	
	Total		10	

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

$$p_1 =$$

$$p_2 =$$

$$p_3 =$$

$$H_S =$$

- 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:

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

[4]

	Pine forest	Warner Mts.	Total	
Haplo_1_00	GGCCTTACCTCCTTACCTCTCCCCCTCCTTTTTTCGCATCCTACTCTTC	1	5	6
Haplo_2_94C.....	7		7
Haplo_3_06C.....A.....T.....	2		2
Haplo_4_14	AAT.CCGTCTTAC..TC.CTTTTT.TTCCC.CCTA.TCTTC..CT.CT		3	3
Haplo_5_73	AATTCGGTCTTACGTTCTCTTTTTT.TTCCCCCCTA.TCTTC.TCTCCT		1	1
Haplo_6_74	AATTCGGTCTTACGTTCTC.TTTTTCTTCCCCC.ATCCTCGTCTCCT		1	1
	Total	10	10	20

Pine forest

Haplo_1_00	GGCCTTACCTCCTTACCTCTCCCCCTCCTTTTTTCGCATCCTACTCTTC	1
Haplo_2_94C.....	7
Haplo_3_06C.....A.....T.....	2
	Total	10

Warner Mts.

Haplo_1_00	GGCCTTACCTCCTTACCTCTCCCCCTCCTTTTTTCGCATCCTACTCTTC	5
Haplo_4_14	AAT.CCGTCTTAC..TC.CTTTTT.TTCCC.CCTA.TCTTC..CT.CT	3
Haplo_5_73	AATTCGGTCTTACGTTCTCTTTTTT.TTCCCCCCTA.TCTTC.TCTCCT	1
Haplo_6_74	AATTCGGTCTTACGTTCTC.TTTTTCTTCCCCC.ATCCTCGTCTCCT	1
	Total	10

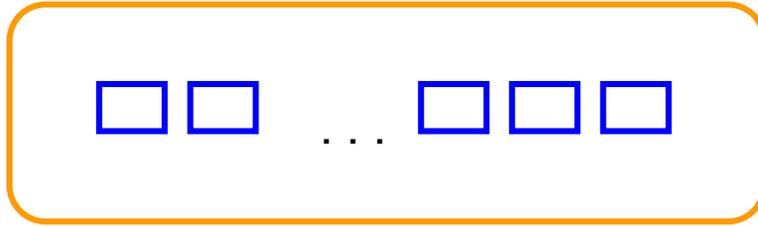
Número de diferencias entre pares de alelos (haplotipos):

	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.

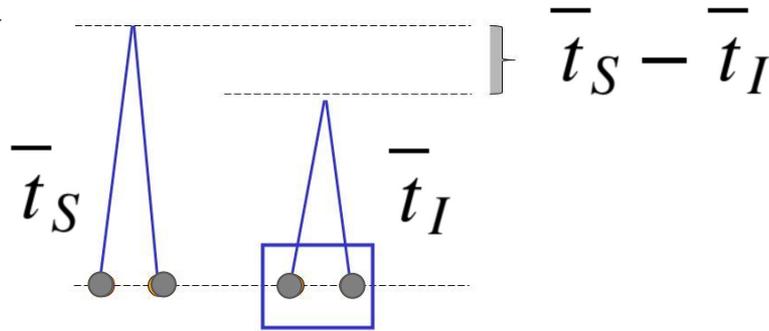
	Pine Forest	Warner Mts.
ADN mitocondrial		
H_S		0.64
π_S		22.74
microsatélites		
H_S	0.37	0.78

Estructura (reproductiva) local : caracterización



Subpoblación
Individuos

$$t_{XY} = \frac{\bar{t}_Y - \bar{t}_X}{\bar{t}_Y}$$



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

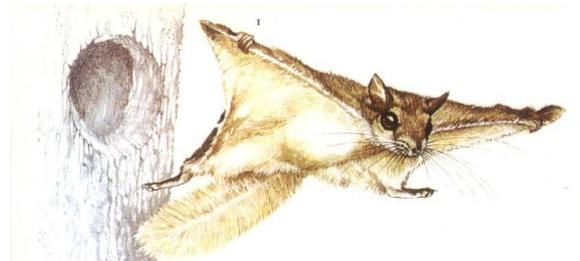
$H_S = H_e$: heterocigosidad esperada (HW)

$H_I = H_o$: heterocigosidad observada en los individuos

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.

Ejemplo: ardillas voladoras (*Glaucomys*) (Alaska)

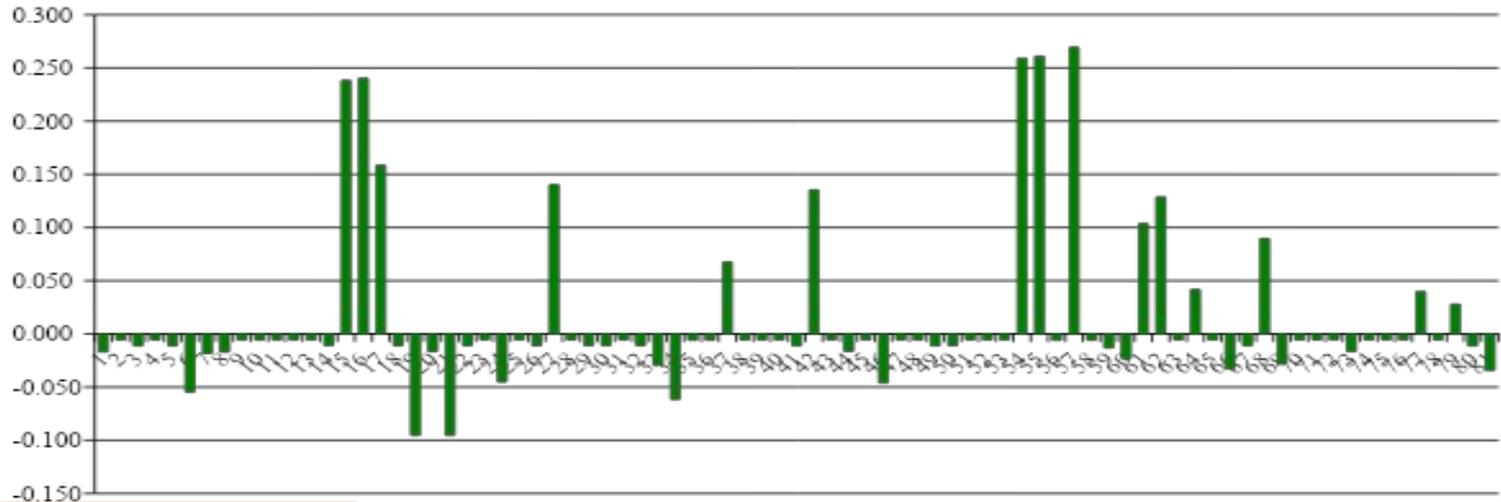
Estimaciones de endogamia local
(5 loci de microsatélites)



Localidad	$H_s = H_e$	$H_i = H_o$	F_{is}
Promedio islas	0.359	0.329	0.074
Prom. continental	0.610	0.612	-0.003

- En continente, interpretamos F_{is} como igual a cero.
- En las islas, deberíamos investigar:
 - incertidumbre de la estimación
 - problemas técnicos (alelos nulos)
 - régimen de apareamientos diferente
 - tamaños poblacionales muy pequeños

Ejemplo: SNPs en un locus humano



SNPs del locus c26b1 humano

Sitio	Genotipo 1	frec(1)	Genotipo 2	frec(2)	Genotipo 3	frec(3)	p	q	p+q	Hs	Hi	Fis
1729	AA	88	AG	3	GG	0	0.984	0.016	1	0.032	0.033	-0.017
1772	CC	90	CG	1	GG	0	0.995	0.005	1	0.011	0.011	-0.006
4492	CC	0	CG	2	GG	89	0.011	0.989	1	0.022	0.022	-0.011
4579	AA	90	AG	1	GG	0	0.995	0.005	1	0.011	0.011	-0.006

Ejemplo: anemia falciforme

Genotipo	Freq Obs Adultos	Frec Esp H-W	$H_e = H_s$	$H_o = H_i$
SS	29	187.4		
SA	2993	2672.4	0.217	0.242
AA	9365	9527.2		
Total	12387	12387		

- Hay un exceso de heterocigotas.
- $F_{IS} = -0,115$ (exceso de heterocigotas de 12%).
- En este caso, la causa del apartamiento no es demográfica (exogamia) sino selectiva (selección a favor de los heterocigotas).
- Por otra parte, las medidas de heterocigosidad no capturan un aspecto importante del caso (gran déficit de homocigotas SS).

Endogamia local (en las subpoblaciones)

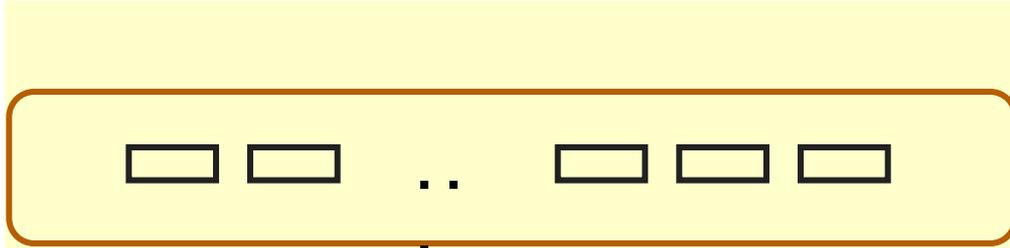
freq Heterocigotas	bajo neutralidad	si hay selección que favorece
$F_{IS} = -1 > H_e$	Exogamia	heterocigotas
$F_{IS} = 0 = H_e$	Panmixia	—
$F_{IS} = 1 < H_e$	Endogamia	homocigotas

Interpretación:

Bajo neutralidad, los apartamientos de lo esperado bajo panmixia deben atribuirse al régimen de apareamiento de los individuos en las subpoblaciones.

La selección puede producir un exceso o un déficit de heterocigotas.

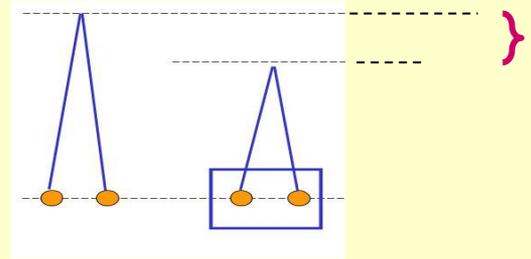
Subdivisión geográfica



$$\bar{F}_{IS} = \frac{\bar{H}_S - \bar{H}_I}{\bar{H}_S}$$

$$\bar{\pi}_{IS} = \frac{\bar{\pi}_S - \bar{\pi}_I}{\bar{\pi}_S}$$

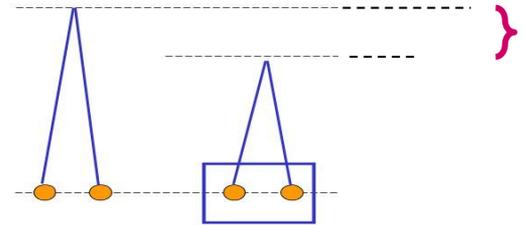
Subpoblación Individuos



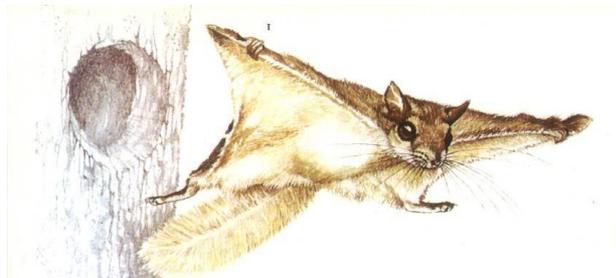
$$\bar{F}_{ST} = \frac{\bar{H}_T - \bar{H}_S}{\bar{H}_T}$$

$$\bar{\pi}_{ST} = \frac{\bar{\pi}_T - \bar{\pi}_S}{\bar{\pi}_T}$$

Población Total Subpoblaciones



Ejemplo: ardillas voladoras (*Glaucomys*) (Alaska)



Estimaciones de endogamia local

Localidad	$H_s = H_e$	$H_i = H_o$	F_{is}
Promedio islas	0.359	0.329	0.074
Prom. continental	0.610	0.612	-0.003

Estimaciones de subdivisión geográfica

Comparación

	F_{ST}
Entre islas del complejo Príncipe de Gales	0.003 - 0.03
Entre islas del complejo y área continental adyacente	0.193
Entre 2 áreas continentales aisladas	0.375

$$-1 < F_{IS} < 1$$

$$0 < F_{ST} < 1$$

$$-1 < F_{IT} < 1$$

freq Heterocigotas	bajo neutralidad	si hay selección favorece
$F_{IS} = -1 > H_e$	Exogamia	heterocigotas
$F_{IS} = 0 = H_e$	Panmixia	—
$F_{IS} = 1 < H_e$	Endogamia	homocigotas

Fst = 0 Subpoblaciones idénticas

Fst = 1 Fijación de alelos alternativos

efectos de la selección y/o de la divergencia neutral

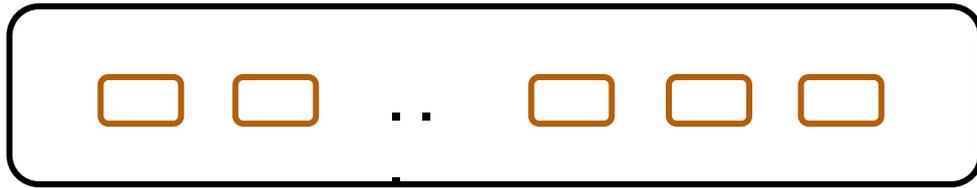
Combinación de efectos intra e interpoblacionales

F_{IS} vs. F_{ST}

F_{IS} : puede ser negativo (exceso de heterocigotas), cero (panmixia) o positivo

F_{ST} : puede ser cero (no hay divergencia geográfica) o positivo

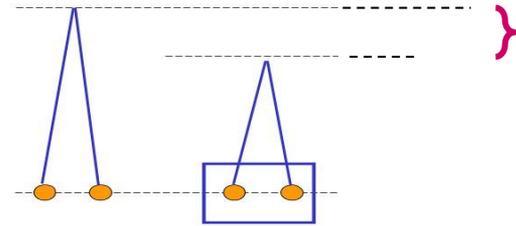
Si la subdivisión es real, el parentesco medio entre alelos dentro de las subpoblaciones será igual o mayor al parentesco medio entre alelos de la población total



$$\bar{F}_{ST} = \frac{\bar{H}_T - \bar{H}_S}{\bar{H}_T}$$

$$\bar{\pi}_{ST} = \frac{\bar{\pi}_T - \bar{\pi}_S}{\bar{\pi}_T}$$

Población Total Subpoblaciones



F_{IS} vs. F_{ST}

- Son independientes: puede haber panmixia dentro de las subpoblaciones y estructura geográfica marcada, o endogamia local sin divergencia geográfica...
- Los genotipos de los individuos solamente afectan F_{IS} (más en general, estadísticos donde aparece la “I”).
 F_{ST} se basa solamente en las frecuencias alélicas, por lo que no se ve afectado por la estructura reproductiva local.