


EFOP-3.4.3-16-2016-00014





Pénzes Zsolt, Markó Bálint

Molekuláris ökológia: populációk szerkezete II.


Jelen tananyag a Szegei Tudományegyetemen
készült az Európai Unió támogatásával.

Projekt azonosító: EFOP-3.4.3-16-2016-00014

Szegei Tudományegyetem
Cím: 6720 Szeged, Dugonics tér 13.
www.u-szeged.hu
www.szechenyi2020.hu

Európai Unió
Európai Szociális
Alap



BEFECTETÉS A JÖVŐBE

A molekuláris ökológia előadások célja a molekuláris módszerek néhány alkalmazási lehetőségének bemutatása ökológiai és evolúcióböölógiai problémák megfogalmazásában/megválaszolásában. Kérdéseink populációkra, fajokra vonatkoznak – például populációk izolációjának mértéke, egy invazív faj eredete, leszármazási kapcsolatok. A válasz keresése során a molekuláris módszerek eszközökként szolgálnak.

Az előadáson a populációk közötti migráció, génáramlás és divergencia koncepcionális modelljeinek tárgyalását folytatjuk, elsősorban a koaleszcencia folyamatot véve alapul.

Ajánlott irodalom: Nielsen R, Slatkin M (2013) An Introduction to Population Genetics. Sinauer. (4. és 5. fejezetek)

Feltételezett előismeretek: molekuláris markerek, a genetikai változatosság mérőszámai, ideális populáció, neutrális változás, koaleszcens modell, populációk szerkezete.

Az olvasólecke átnézése kb. 20 percet vesz igénybe.

Migráció és F_{ST}

- Végtelen pozíció modell – két szekvencia ($n = 2$)
 - páronkénti különbségek várható száma: $E[t]\theta$
 - minta egy populációból ($E_S[t]\theta$): 2θ
 - minta két különböző populációból ($E_D[t]\theta$): $(1/(2M) + 2)\theta$
- F_{ST} a végtelen pozíció modellben
 - k szekvencia pozíció ($k \rightarrow \infty$, így $\theta/k \rightarrow 0$)
 - heterozigotizáció/pozíció \equiv páronkénti különbségek száma/pozíció
 - egy populáció mintázása:

$$H_S = 2\theta/k$$

- mintázás egy vagy két populációból azonos valószínűséggel, így

$$H_T = (2\theta/k + [1/(2M) + 2]\theta/k)/2 = [1/(4M) + 2]\theta/k$$

- F_{ST} két populációra:

$$F_{ST} = \frac{H_T - H_S}{H_T} = 1 - \frac{2\theta/k}{[1/(4M) + 2]\theta/k} = \frac{1}{1 + 8M}$$

Értelmezzük a populációk szerkezetének jellemzésére használt F_{ST} mérőszámot a migrációval kiegészített standard koaleszcencia folyamattal. Ilyen módon a szerkezet molekuláris markerekkel történő becslésével a migráció folyamatára következtethetünk, közvetlenül összekapcsolva a marker, például DNS szekvencia adatokat és a migrációs rátát. **DNS szekvenciák** végtelen pozíció modelljét feltételezve beláttuk, hogy a páronkénti különbségek várható száma $E[t]\theta$, ahol t a koaleszcencia idő. A két populációra bevezetett modellre alkalmazva ($n = 2$, két azonos méretű populáció és szimmetrikus migráció feltételekkel), a páronkénti különbségek várható száma kifejezhető (a koaleszcencia idő várható értéke $E_S[t] = 2$ ha a minta egy, $E_D[t] = 1/(2M) + 2$ ha két különböző populációból származik).

A végtelen pozíció modellben nagy számú szekvencia pozíciót (k) tételezünk fel (pontosabban $k \rightarrow \infty$, így a pozíciónkénti egynél több mutáció esélye elhanyagolható), ennek következtében az egy pozícióra vonatkoztatott páronkénti különbség szám várható értéke és az egy pozícióra vonatkoztatott várható heterozigotizáció ekvivalens. Ez alapján kifejezhetőek a heterozigotizáció értékek és ezek ismeretében az F_{ST} (a várható heterozigotizáció $E_S[t]\theta/k$ ha egy, $E_D[t]\theta/k$ ha két különböző populációból származik a minta).

Koaleszcencia migrációval – F_{ST}

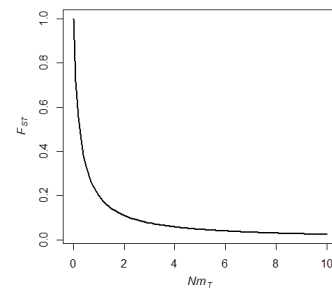
- Kiterjeszhető a különböző modellekre – pl. sziget modell:

$$F_{ST} = \frac{(d-1)/d}{(d-1)/d + 2dM}$$

- egy populációba bevándorlók száma generációnként: $m_T = (d-1)m$
- sok populáció ($d \rightarrow \infty$) és $M = 2Nm$, ekkor

$$F_{ST} = \frac{1}{1 + 4Nm_T}$$

- F_{ST} gyorsan csökken Nm_T növekedésével
- gyakorlatban sokszor $Nm_T > 1$ esetén egy populációnak tekintik
- de durva tapasztalati szabály – pl. $Nm_T = 10$ esetén $F_{ST} \approx 0,024$

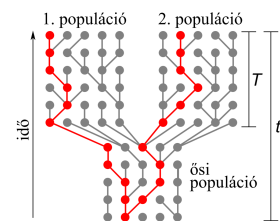


Két azonos méretű populációra $F_{ST} = 1/(1 + 8M)$. Az eredmény kiterjeszhető további modellekre, így a korábban tárgyalt sziget modellre is, ahol d a populációk száma. Szimmetrikus migrációt és minden páronkénti kombinációban azonos migrációs rátát feltételezve a teljes migrációs ráta $m_T = (d-1)m$ egy adott populációba generációnként ($d-1$ populációból történhet migráció, mindegyik esetre m rátával). Nagy populáció számot feltételezve ($d \rightarrow \infty$), egy széles körben alkalmazott összefüggést kapunk: $F_{ST} = 1/(1 + 4Nm_T)$. A modell feltételei által meghatározott körülmények között az F_{ST} ismeretében a bevándorló egyedek számára következtethetünk. Az eredmény azonos S. Wright más módon és kissé más feltételekkel kapott összefüggésével.

Az ábra az F_{ST} bevándorló egyedek számától (Nm_T) való függését szemlélteti a sziget modell alapján. Az F_{ST} gyorsan csökken Nm_T növekedésével. A gyakorlatban sokszor 1 feletti Nm_T érték esetén már egy populációról beszélnek, azonban ezt az ábra alapján óvatosabban kell kezelni.

Divergencia modellek

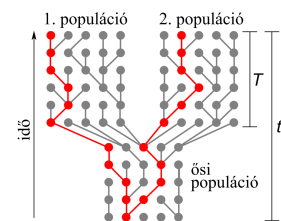
- Populáció szerkezet sziget modelljeinek tipikus feltételei
 - elkülönülés régen, ma egyensúly
 - folyamatos génáramlás
- Gyakran valótlan feltételezés
 - pl. *Homo sapiens* out-of-Africa (60-110 ezer éve)
 - pl. hosszúszárnnyú bálna – ma alapvetően izoláció
- A populáció szerkezet divergencia modelljei
 - divergencia egy közös ősből
 - ezt követően nincs génáramlás (izoláció)
 - ma független WF populációkként evolválódnak
- koaleszcencia folyamat két divergáló populációval
- minta (szekvencia) két populációból
- populációk szétválása a T időpontban (jelentő, divergencia idő)
- MRCA az ősi populációban (koaleszcencia, t)



A kontinens-sziget vagy sziget modelleken túl további modell típusokat is használnak a gyakorlatban a szerkezet figyelembe vételére. Közös jellemzőjük, hogy feltételezik azt, hogy a különálló populációkra, al-populációkra tagolódás már régen megtörtént, beállt az egyensúlyi állapot és folyamatos a génáramlás. Sok esetben azonban ezek a feltételek nem realiztikusak. A *Homo sapiens* elterjedésére nem alkalmazható egyik említett modell sem, az Afrikából történő elterjedés sokkal inkább kis populációkkal történő kolonizáció sorozattal magyarázható korlátozott génáramlással, erre különböző jellegű adatok utalnak. Génáramlás hiányában populációk közötti különbség fokozatosan nő (pl. mutáció és drift hatására), ezt nevezük populáció divergenciának. A **divergencia modellek** olyan esetekre alkalmazhatóak, ahol divergencia történik egy ősi populációból és közben nincs génáramlás a populációk között. Az ábra a koaleszcencia folyamatot szemlélteti két divergáló populációban, ahol a populációk szétválása T idővel ezelőtt történt (**divergencia idő**). A **minta** utolsó közös őse (MRCA) az ősi populációban van, vagyis a koaleszcencia idő (t) nagyobb, mint a divergencia idő.

Divergencia modellek

- Gyakran további feltételek
 - pl. humán demográfia – palacknyak a divergencia kezdetén
 - tfh. a populációk mérete azonos (ősi: N_A), $N_A = N_1 = N_2 = N$
- Divergencia és koaleszcencia idő nem azonos ($t \neq T$)
 - minták két populációból (szekvenca) – de koaleszcencia idővel a divergencia idő nem becsülhető (gyakori hiba)
- Koaleszcencia idő ($n = 2$ eset, idő $2N$ generációban mérve)
 - minta egy populációból: $E_S[t] = 1$ (standard koaleszcencia modell)
 - minta különböző populációból: $E_D[t] = T + 1$ (populációk egyesülése után standard koaleszcencia modell)
- minta két populációból ($n = 2$)
- ősök követése a két populációban egymástól függetlenül $2NT$ generáción keresztül
- majd az ősök populációinak egyesítése – divergencia idő: T
- és követése az MRCA-ig – koaleszcencia idő: t



Gyakran további feltételekkel élünk a divergencia kezdetére, populációk demográfiájára. Például gyakori feltételezés a humán modellekre a palacknyak hatás a divergencia kezdetén. Az egyszerűség kedvéért feltételezzük a továbbiakban, hogy a populációk mérete azonos és nem változik.

A koaleszcencia idő várható értékének meghatározása az eddigiekhez hasonlóan történik. Az ábra olyan esetet szemléltet, ahol a **minta** két különböző populációból származik. Ez esetben a jelenből kiindulva követjük az őseket a divergencia eseményig (divergencia idő), majd az egyesített populációban az utolsó közös ősig (koaleszcencia).

A fenti egyszerűsítések mellett két azonos populációból származó mintára (nincs migráció, így minden ő a kérdéses populációban vagy az ősi populációban van) a standard koaleszcencia modellt alkalmazhatjuk, vagyis $E_S[t] = 1$ az időt $2N$ generációban mérve ($n = 2$ eset). Ha a minta nem egy populációból származik, a koaleszcencia a populációk egyesülése után következhet csak be, az ősi populációban. Ez utóbbira már a standard koaleszcencia modellt alkalmazzuk, így $E_D[t] = T + 1$.

A modellek kiterjeszthetőek egyéb esetekre is, például az ősi populáció mérete (N_A) eltérhet, vagy a populációk mérete változhat.

Divergencia modellek

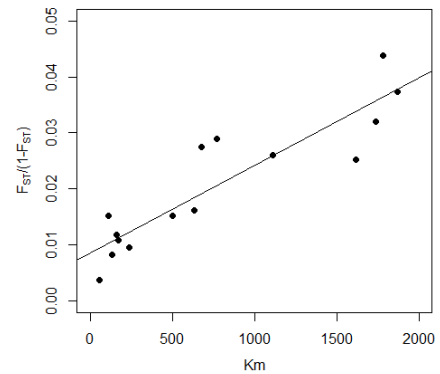
- Végtelen pozíció modell – várható páronkénti különbségek száma:
 - θ ha a szekvenciák egy, $\theta(T + 1)$ ha eltérő populációkból
 - így $H_S = \theta/k$ és $H_T = (T/2 + 1)\theta/k$, illetve
$$F_{ST} = 1 - \frac{H_S}{H_T} = \frac{T}{T + 2}$$
 - $T = 0$ esetén $F_{ST} = 0$ – homogén populáció
 - nagy divergencia idő esetén $F_{ST} \approx 1$
 - de erős feltételek a populáció méretre
- Adott szerkezet mérőszám, pl. F_{ST} – következtetés a populáció történetre?
 - divergencia és migrációs modellekkel is értelmezhető – F_{ST} alapján nem tudunk különbséget tenni
 - F_{ST} nem informatív a modellek alkalmazhatóságára

Ha a mintáink szekvenciák, a végtelen pozíció modellt feltételezve a páronkénti különbségek várható értéke és ebből az egy pozícióra vonatkoztatott heterozigotizáció a migrációs modellekhez hasonló módon becsülhető. Mindezek alapján az F_{ST} érték és a divergencia idő kapcsolatára (az egyszerűsítő feltételek mellett) azt kapjuk, hogy $F_{ST} = T/(T + 2)$.

Ha $T = 0$, nincs szerkezet, így $F_{ST} = 0$. A divergencia idő növekedésével F_{ST} értéke 1-hez közelít. Azonban szigorú megkötésekkel élünk a populációk méretére vonatkozóan, így a fenti összefüggés az F_{ST} -re kevésbé informatív, összevetve a migrációs modellekkel. Mindamellettt beláttuk, hogy egy adott F_{ST} érték magyarázható migrációs modellekkel és divergencia modellekkel egyaránt. Pusztán egy pillanatnyi helyzet alapján becsült szerkezeti mérőszám alapján nem tudunk a populáció történeti modellek között dönteni, ehhez további információra van szükség.

Populáció szerkezet és földrajzi távolság

- Differenciálódás mértéke (F_{ST}) a földrajzi (fizikai) távolsággal gyakran nő – „isolation by distance”
- kaliforniai oroszlánfóka (*Zalophus californianus*) – Kaliforniai-öböl, populáció párok
- mtDNS szekvencia és mikroszatellitek
- $F_{ST}/(1 - F_{ST})$ megközelítőleg lineáris kapcsolata a földrajzi távolsággal – linearizált F_{ST} , lineáris regresszió



A populációk differenciálódásának mértéke (például populáció páronként számolt F_{ST} értékkel jellemezve) gyakran nő a populációk földrajzi távolságával, ez esetben **„isolation by distance”** (izoláció a távolság alapján) mintázatról beszélünk. Az ábra egy ezt szemléltető esettanulmány eredményét mutatja, ahol egy pont két populáció genetikai (differenciálódás mértéke) és kilométerben kifejezett földrajzi távolságának felel meg.

Az ábrán azonban nem a **páronkénti** F_{ST} értékét, hanem annak un. linearizált formáját, $F_{ST}/(1 - F_{ST})$, látjuk. Ebben az átalakított formában a fizikai távolsággal hozzávetőleg lineáris összefüggést mutat.

Számos példát találunk ilyen összefüggésre a humán populációk viszonylatában is. A humán evolúció „out of Africa” hipotézisét sok adat támasztja alá, egy tipikus „isolation by distance” mintázatot eredményez. A szekvenciális kolonizációs események sorozata nagy valószínűséggel fontos eleme lehet a magyarázatának, az adatokkal a modellek konzisztens eredményt adnak. De hogyan értelmezhető ez a lineáris összefüggés a modellek alapján?

Populáció szerkezet és földrajzi távolság

- „Isolation by distance” magyarázata a modellek alapján
 - egyszerűsítő feltételek mellett két populációra $F_{ST} = 1/(1 + 8M)$
 - így $M = (1 - F_{ST})/(8F_{ST})$
 - lineáris összefüggés, ha a migrációs ráta a földrajzi távolság lineáris függvénye
- Ha migráció csak a szomszédos populációk között – „stepping stone” modellek
 - speciális eset, M a populációk távolságának valamilyen függvénye
 - célja általában az „isolation by distance” mintázat értelmezése
- Korrelációt a divergencia modellek is magyarázhatják
 - pl. divergencia események sorozata, új populáció a szomszédos területen
 - szomszédos területek genetikai hasonlósága – a távolságfüggő migrációs modellekhez hasonlóan
 - $F_{ST} = T/(T + 2)$, így $F_{ST}/(1 - F_{ST})$ lineáris kapcsolata a földrajzi távolsággal, ha a populációk divergenciája a divergencia idő (T) lineáris függvénye

Korábban láttuk, hogy adott feltételek mellett $F_{ST} = 1/(1 + 8M)$, amely a fenti módon linearizálható. Vagyis ha M migrációs ráta a fizikai távolság **lineáris függvénye**, akkor $F_{ST}/(1 - F_{ST})$ is az. Ez magyarázhatja például a kaliforniai oroszlánfókára kapott eredményeket: pl. migráció bármely két populáció között előfordulhat, azonban ennek esélye a távolsággal csökken. A **„stepping stone”** modell egy speciális eset, ahol migráció csak a szomszédos populációba történhet.

De hasonló mintázatot eredményezhetnek a divergencia modellek is. Például ha divergencia események sorozata követi egymást és az új populáció a szomszédos területeket kolonizálja. Azt várjuk, hogy a szomszédos területek populációi között nagyobb a hasonlóság, kisebb mértékű a genetikai differenciáció, mint a távolabbiak között. Ez utóbbi esetben több divergencia esemény történt, több idő telt el szétválásuk óta. Ha a divergencia mértéke T lineáris függvénye, az eredmény hasonló a migrációs modellekkel kapotthoz.

Az „isolation by distance” mintázat magyarázható migrációs és divergencia modellekkel egyaránt. Viták tárgyát képezi például, hogy a divergencia modellek, folyamatos génáramlás modellek, vagy a kettő keveréke magyarázza legjobban a *Homo sapiens* genetikai változottság mintázatát.

Következtetés a gyakorlatban

- Molekuláris markerek elemzése → populáció méret becslése, migráció és divergencia mintázatok...
- Elemzés: elméleti modellek → módszerek a modellek tesztelésére, a modellek paramétereinek becslésére
 - alapvető paraméterek pl. θ , M , T – következtetés a paraméterekre minták alapján
 - pl. π , $S \rightarrow \theta$ becslése
 - pl. H becslése (H_S , H_T) → $F_{ST} \rightarrow M$ és T
 - különböző modellek és becslési módszerek
- Példa: óriás panda (*Ailuropoda melanoleuca*)
 - elterjedése: Kína, foltokban – populációk közötti génáramlás?
 - minta: 2 populáció, $n = 63$, 13 mikroszatellit lokusz
 - $F_{ST} = 0,26$
 - sziget modell alapján $M = (1 - 0,26)/(8 * 0,26) = 0,36$
 - 0,36 migráns a populációk között generációnként – egyik kihalásával a genetikai változatosság jelentős csökkenése

Populációk genetikai adataiból következtethetünk a populáció méretére (demográfiára), populációk közötti génáramlásra, divergenciára (a populációk történetre) és még számos további jellemzőre. Ehhez különböző módszerek állnak rendelkezésre. A koaleszcencia elmélet különösen hasznosnak bizonyult, hiszen a mintákra fókuszál.

Különböző populáció **paraméterekkel** találkoztunk, ilyenek a θ ($= 4N\mu$), a migrációs ráta ($M = 2Nm$) vagy a divergencia idő (T). Ezeket az adatokból (mintákból) számolt **statisztikákkal** becsüljük, ilyen volt a páronkénti különbségek átlagos száma (π) vagy a szegregálódó pozíciók száma (S) szekvenciákra (θ becslésére). Al- lél és genotípus adatokból becsüljük a populáció heterozigotizációját, több populáció esetén a szerkezetet figyelembe véve (H_S) vagy figyelmen kívül hagyva (H_T). Ezt felhasználhatjuk az F_{ST} becslésére ($F_{ST} = 1 - H_S/H_T$) és ez alapján következtethetünk a migrációs rátára ($M = (1 - F_{ST})/(8F_{ST})$) vagy a divergencia időre ($T = 2F_{ST}/(1 - F_{ST})$). M és T becslésére különböző módszerek léteznek, pl. a különböző modelleknek megfelelően.

Az óriás panda populációk közötti génáramlásra kapott eredmények arra utalnak, hogy több populáció genetikailag jelentősen elkülönült, így az egyik kihalása esetén a genetikai változatosság jelentős mértékben lecsökken. A becsléshez a sziget modellt használták.

Ellenőrző kérdések

1. Hogyan változik az F_{ST} a migrációs ráta növekedésével?
2. Hogyan változik az F_{ST} a divergencia idő növekedésével?
3. Milyen feltételei vannak a divergencia modelleknek?
4. Mi a különbség a koaleszcencia idő és divergencia idő között?
5. Mit értünk „isolation by distance” mintázat alatt?
6. Milyen feltételei vannak a „stepping stone” modellnek?
7. Mikor várunk lineáris kapcsolatot a földrajzi távolság és az $F_{ST}/(1 - F_{ST})$ alakban linearizált F_{ST} között?
8. Sorolja fel, hogy milyen logikai lépéseken keresztül tudjuk szekvencia adatokból az F_{ST} értékét becsülni!



EFOP-3.4.3-16-2016-00014



Jelen tananyag a Szegedi Tudományegyetemen készült az Európai Unió támogatásával. Projekt azonosító: EFOP-3.4.3-16-2016-00014

Szegedi Tudományegyetem
Cím: 6720 Szeged, Dugonics tér 13.
www.u-szeged.hu
www.szechenyi2020.hu



MAGYARORSZÁG
KORMÁNYA

Európai Unió
Európai Szociális
Alap



BEFEKTETÉS A JÖVŐBE