

EFOP-3.4.3-16-2016-00014

MOLEKULÁRIS ÖKOLÓGIA: A GENETIKAI VÁLTOZATOSSÁG MÉRŐSZÁMAI

AP4_TTIK Kárpát-medencei oktatási tér
kialakítása**PÉNZES ZSOLT
MARKÓ BÁLINT**

SZÉCHENYI 2020

MAGYARORSZÁG
KORMÁNYAEurópai Unió
Európai Szociális
Alap

BEFEKTETÉS A JÖVŐBE

A molekuláris ökológia előadások célja a molekuláris módszerek néhány alkalmazási lehetőségének bemutatása ökológiai és evolúcióböiológiai 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 populáció genetikai változatosságának jellemzésére használt legfontosabb mérőszámokat tárgyaljuk.

Genetikai változatosság – populáció

A genetikai változatosság értelmezése egy lokuszon

- Populáció génkészlete
 N diploid egyed: $2N$ db allél egy adott lokuszon
- Két allél típus (állapot, A és a) esetén
 AA , Aa , aa genotípusok
- A genotípusok relatív gyakorisága (genotípusos változatosság):
 - AA , ill. aa homozigóták: $f_{AA} = N_{AA}/N$, $f_{aa} = N_{aa}/N$
 - Aa heterozigóta (H): $f_{Aa} = N_{Aa}/N$
 - $N_{AA} + N_{Aa} + N_{aa} = N$, és $f_{AA} + f_{Aa} + f_{aa} = 1$
- Allélok relatív gyakorisága (f_A és f_a , $f_A + f_a = 1$):

$$f_A = \frac{2 * N_{AA} + N_{Aa}}{2N} = f_{AA} + \frac{f_{Aa}}{2} = 1 - f_a$$

$$f_a = \frac{2 * N_{aa} + N_{Aa}}{2N} = f_{aa} + \frac{f_{Aa}}{2} = 1 - f_A$$

Célunk a **genetikai változatosság**, az allélok és genotípusok gyakoriságának (allél és genotípus eloszlásnak) becslése egy adott populációban egy tetszőleges lokuszra – például mikroszatellitire, szekvencia pozícióra.

Ha mindössze egy allél típus (allél állapot, pl. A) található a populációban a kérdéses lokuszra (így minden egyed AA homozigóta), akkor nincs változatosság, a populációt **monomorf**nek nevezzük. Genetikai változatosságról akkor beszélünk, ha minimum két allélt tudunk kimutatni (megfelelő gyakorisággal, lásd polimorfizmus). Két allél esetén összesen három különböző genotípus fordulhat elő a populációban.

Az allél gyakoriság a genotípusok ismeretében számolható: a kérdéses allélra nézve homozigóta egyed 2, a heterozigóta 1, a többi 0 allélt hordoz egy diploid populációban. Viszont allél gyakoriságból genotípus gyakoriság csak kivételes esetben becsülhető. Ehhez az **allél kombinálódás** szabályszerűségeit is ismernünk kell (mint például teljes önmegtermékenyítés esetén vagy egy ideális populációban).

Genetikai változatosság – populáció

- Mendeli karakter: fenotípus – genotípus – allél
 - a tulajdonság mendeli öröklődése
 - a tulajdonság változatossága csak a genotípusos változatosságtól függ, a genotípus egyértelműen meghatározható
 - genotípusból az allél gyakoriság egyértelmű – allél gyakoriságból genotípus gyakoriság csak kivételes esetben
- m db függetlenül szegregálódó allél egy lokuszon:
 - m homozigóta típus
 - $\binom{m}{2} = m(m-1)/2$ heterozigóta típus
- Két lokusz, lokuszonként két allél:
 - AB, Ab, aB, ab ivarsejt típusok („haplotípus” – szekvenciák)
 - $AB/AB, Ab/AB, \dots, ab/ab$ genotípusok (4^2 kombináció)
 - ... amely $3^2 + 1$ genotípus, ha az eredettől eltekintünk:
 $AB/AB, AB/Ab, Ab/Ab,$
 $aB/aB, aB/ab, ab/ab,$
 $AB/aB, Ab/ab,$
 $AB/ab, Ab/aB$ – kétféle (cisz és transz) kettős heterozigóta

Mendeli karakter a tulajdonságok legegyszerűbb típusa, változatossága nem függ a környezet változatosságától és mendeli öröklődést mutat. Az *MC1R* tárgyalt változatosságát eredményező szekvencia pozíció, vagy a mikroszatellitek is mendeli karakterek. Allélnak (karakter állapotnak) a nukleotidot illetve a szekvencia szakasz hosszát tekintjük. Míg *MC1R* esetén maximum két allél volt a populációban, a mikroszatellitek esetén jóval több lehet. Allélszámtól függetlenül a gyakoriságokat a két allélos esethez hasonlóan becsüljük.

Több lokuszt vizsgálva szükség lehet az allélok lokuszok közötti asszociációjának (ivarsejt fázisnak) ismeretére. Ezeket a populációban előforduló lokuszok közötti allél kombinációkat **ivarsejt típusoknak** nevezzük, amely lényegében az allél többlokuszos megfelelője. Egy diploid egyed két „haploid” ivarsejt típussal rendelkezik a kérdéses lokuszokra. A lokuszokat külön értékelve azonban a **gaméta fázisról, genetikai kapcsoltságról** nincs információ, vagyis pl. a kétféle heterozigóta a legtöbb egylokuszos módszerrel nem különíthető el. Szekvenciákra alkalmazva az ivarsejt típust **haplotípusnak** nevezzük, ami így egy egyedi szekvenciát jelent a populációban (a lokusz egy szekvencia pozíció ebben az értelmezési módban). A populáció változatosságának jellemzéséhez az ivarsejt típusok/haplotípusok gyakoriságának ismerete is hozzátartozik.

Genetikai változatosság jellemzése – áttekintés

- Markerek elemzése – mintavétel a populációból (egyedek) és a genomból (lokuszok)
- Mérészámok – változatosság mértékétől, módszerektől, modellektől, statisztikai sajátosságoktól függően
- A fajon belüli genetikai variabilitás szerveződése
 - populáción belül (egyedek között, egy vagy több lokuszon)
 - multilokuszos mintázat – közelítések, többváltozós módszerek
 - heterozigotizáció (heterozigotia) és polimorfizmus
 - DNS szekvencia: pozíciók szintjén, pl. nukleotid diverzitás
 - lokuszok kapcsolata – a kapcsoltsági egyensúly
 - populációk között: fixációs index változatok
- Genetikai (evolúciós) távolságok – a divergencia mértéke
- Kvantitatív (mennyiségi) jellegek: heritabilitás, additív genetikai variancia

Allél, genotípus és ivarsejt típus gyakoriságok hordozzák a **teljes információt** egy adott populáció genetikai változatosságáról. A generációról generációra történő (evolúciós) változásra vonatkozó következtetésekhez több generációról is szükségünk lehet ezekre az alapadatokra.

Azonban különösen ha sok allélunk van, ami például mikroszatellit lokuszok esetén gyakran előfordul, a gyakoriság értékek önmagukban kevésbé szemléletesek, nehezen kezelhetők. Másrészt kérdésünktől függően sokszor részinformáció is elegendő, vagy csak a populáció(k) adott sajátosságait szeretnénk hangsúlyozni. Ezért különböző szempontok alapján **mérészámokat** képezhetünk az alapadatokból. Egy részük elméletileg is megalapozott, ezeket részesítjük előnyben.

Különböző mérőszámaink vannak az egyed feletti **hierarchikus szerveződés** szintjeinek jellemzésére. Vonakozhatnak egy vagy több lokuszra. Hangsúlyozhatják az egy populáción belüli allél gyakoriság különbségeket, másokkal a populációkra történő tagolódást figyelembe véve a populációk közötti különbségeket emeljük ki. A lokuszok közötti allél asszociációt is egy külön mérőszámmal jellemezzük. Más természetű mérőszámaink vannak a kvantitatív jellegek változatosságára. Ez utóbbiakat az elméletünkkel együtt fogjuk tárgyalni.

Genetikai változatosság jellemzése – mérőszámok

- **Heterozigotitás** vagy (Nei-féle) gén diverzitás (jele h vagy H):

$$h = 1 - \sum_{i=1}^m f_i^2$$

- Annak a valószínűsége, hogy a populációban két véletlenszerűen kiválasztott allél különbözik
- Több (L) lokuszra: átlagos heterozigotitás (heterozigotia, $h_L, \bar{H}$):

$$h_L = \frac{1}{L} \sum_l h_l$$

- Informatív, ha az allélok száma nem túl nagy
- Tetszőleges lokuszra becsülhető

A **heterozigotitás** a populáció genetikai változatosságának egyik leggyakrabban használt mérőszáma. Több lokuszra az egyedi lokuszok értékeinek átlagát adjuk meg. A gyakorlatban a populációból vett minta alapján becsüljük. Jelentős változatosság esetén értéke 1-hez közelít, monomorf populációra $H = 0$.

A heterozigotitással tetszőleges lokuszon/lokuszokon számszerűsíthetjük a populáció változatosságát. **Markerek** alkalmazásával azonban célunk az általánosítás például az azt hordozó élőlények populációjára („objektumokra”). Mivel jelentős az eltérés a különböző genom régiók által mutatott változatosságban, ezért

- az általánosítás nagy lokusz szám és/vagy nagy genom régiók elemzését igényli,
- az összehasonlítások, történeti rekonstrukciók ortológokon alapulnak.

Genetikai változatosság jellemzése – mérőszámok

- Alkalmazása: leíró mérőszám
- De elméletileg megalapozott, egyértelmű biológiai jelentés – heterozigóták aránya egy ideális populációban
- Akár információ az evolúciós változásra is speciális feltételek mellett
 - feltételek (Wright-Fisher modell): diploid eset, konstans effektív populációméret (N_e), neutralitás, u mutációs ráta
 - várható értéke:

$$E(h) = \frac{\theta}{1 + \theta}, \text{ ahol } \theta = 4N_e u$$
- θ alapvető paraméter
 - becslése egy egyensúlyi populációban markerekkel $\rightarrow h \rightarrow N_e u$
 - θ becslésére egyéb módszerek is léteznek

Egy ideális diploid populációban a **heterozigozitás** a két különböző allélt hordozó heterozigóták gyakorisága. Definíciónk azonban lehetővé teszi, hogy ne csak diploid esetre alkalmazzuk, így felhasználható például mtDNS marker alapú következtetésben is, mivel a mtDNS-t haploid állapotnak tekintjük. Nem csak allél állapotra, hanem allél eredetre is kiterjeszthető, ez esetben a heterozigozitás annak a valószínűsége, hogy a két kiválasztott allél különböző eredetű (eltérő ősi allélből származik).

Sőt, magára a folyamatra is utalhat neutrális esetben (nem hat természetes szelekció). A θ (théta) a genetikai változatosság egy központi jelentőségű paramétere.

Természetesen a heterozigozitást akkor is használhatjuk, ha a fenti **modellek** (ideális populáció, Wright-Fisher modell) feltételei nem állnak fenn (nem ideális populáció, szelekció hat, nincs egyensúly stb.). Ez esetben mint leíró mérőszámot használjuk például populációk összehasonlítása során. A **tapasztalt** heterozigozitás eltérése az ideális populációban **várt** értéktől a populáció számos sajátosságra utalhat, tesztelése gyakran az első lépés egy populáció genetikai változatosságának vizsgálata során.

Genetikai változatosság jellemzése – mérőszámok

- **Polimorfizmus:** a polimorf lokuszok aránya
- Polimorf lokusz – szubjektív kritikus értékek, például
 - allél gyakoriság $< 99\%$,
 - 95% és egy további allél gyakorisága $> 1\%$.
- Kis mértékű polimorfizmus esetén informatív
- Nagy mintákra, sok lokuszra – becslése kis mintából?
- Példa: Anglia, európai populáció (Harris 1972)
 - enzimpolimorfizmus, 71 lokusz
 - 51 monomorf, 20 polimorf $\rightarrow$ polimorfizmus: $20/71 = 0.282$
 - átlagos heterozigotizáció: 0.067

A **polimorfizmust** a változatosság szinonimájaként is használjuk a gyakorlatban, azonban mint a genetikai változatosság mérőszámának jelentése rögzített: a polimorf lokuszok arányát jelenti. Azonban az, hogy mikor tekintünk egy lokuszt polimorfnak már kevésbé egyértelmű. Tipikus feltétel (lásd pl. az SNP definícióját), hogy legalább két alléljának gyakorisága 1% feletti a populációban. Azonban az 1% körüli gyakoriságú allélok kimutatásához nagy mintára van szükség, vagyis a megbízható becsléshez nagy mintanagyságok kellenek. Másrészt az 1% egy önkényes küszöbérték (vagyis elméleti szempontból nem megalapozott), így nem meglepő, hogy önmagában nem informatív az evolúciós háttér folyamatokra.

Harris és mtsai. 71 **enzim lokusz** alléljait vizsgálták egy populációban. Azt találták, hogy ebből 20 volt polimorf, akár kettőnél több alléllal a populációban, ami 28%-os polimorfizmust jelent. A heterozigotizációt lokuszonként megbecsülve és a lokuszokra kapott értéket átlagolva ez közel 7%-nak adódott. A becslés a hagyományos enzimpolimorfizmus vizsgálattal történt, markerekként használva a lokuszokat (hiszen kérdésünk nem az egyedi enzimekre vonatkozott).

Genetikai változatosság jellemzése – mérőszámok

- Jelentős különbségek taxonok között a változatosság mértékében
- Sok „kivételes” eset – a mechanizmus ismeretében értelmezhető
 - északi elefántfóka, gepárd – nincs kimutatott genetikai változatosság
 - árpa – jelentős polimorfizmus de alacsony heterozigotizás
- Enzimpolimorfizmus eredmények (vizsgált lokusz szám, polimorf lokusz arány és átlagos heterozigotizás):

Élőlény	Lokusz	Polimorfizmus	Heterozigotizás
Északi elefántfóka	24	0	0
Elefánt	32	0.29	0.089
Gepárd	32	0	0
<i>Drosophila pseudoobscura</i>	24	0.42	0.12
Árpa	28	0.30	0.003

Nagyobb heterozigotizás általában nagyobb polimorf lokusz arányt is jelent.

Azonban **élőlény csoportok között** is jelentős eltérések lehetnek. Gerinctelenekre általában a változatosság gyakran nagyobb, mint a gerincesek esetén. De például az északi elefántfóka esetén gyakorlatilag nincs enzimpolimorfizmussal kimutatható változatosság, szekvenciákkal is minimális, összevetve például a déli elefántfókával. Ez a különbség speciális populáció sajátosságokra utalhat. Az északi elefántfóka vagy a gepárd esetén a kis populációméret következtében felerősödő genetikai sodródás egy valószínű magyarázat. Az árpa esetén tapasztalt mintázat, a vártnál alacsonyabb átlagos heterozigotizás, az öntermékenyítés következménye. Vagyis a genetikai változatosság mintázat különbségeket a változatosságot formáló folyamatokkal, életmenet jellemzőkkel magyaráztuk.

Mivel az egyes **genom régiók** változatosságában jelentős különbségek lehetnek, ezért a következtetéshez több lokusz vizsgálata szükséges.

Genetikai változatosság jellemzése – DNS szekvencia

- Populáció genetikai változatossága – DNS szekvenciák, számos előny (elmélet, módszerek)...
- SNP: polimorf szekvencia pozíciók – gyakoriság
- Haplotípus: egyedi szekvencia – gyakoriság és genetikai távolság
- **Szegregálódó pozíciók száma (S)** a szekvenciában
 - világos biológiai jelentés – ha karakterenként egy változás (végtelen allél modell): közös őstől számolva a mutációk száma
 - folyamat (Wright-Fisher modell, S normalizált értékére):

$$E(\tilde{S}) = \theta$$

A DNS szekvenciákat több különböző módon is elemezhetjük. Egyrészt történhet **pozíciónként**, a 4 lehetséges állapot gyakoriságát becsüljük (nukleotid, lásd SNP).

Tekinthetünk egy genom régiót lokusznak, alléljai a **szekvenciák** (haplotípusok). Amennyiben csak szekvencia azonossággal (vagyis haplotípus gyakorisággal) számolunk, az elemzés a mikroszatellitekhez hasonlóan történik.

Azonban a szekvenciák közötti különbség számszerűsíthető, közöttük **genetikai távolság** definiálható. Nagy távolság nagyobb mértékű divergenciára utal. Érdemes megjegyezni, hogy amennyiben indel események következtek be (így a szekvenciák hossza eltér), elemzés előtt a szekvenciák illesztésére lehet szükség. Ennek tárgyalásától ezen a kurzuson eltekintünk.

A szekvencia szintű változatosság jellemzésének legegyszerűbb módja a **szegregálódó pozíció szám**. Amennyiben egy pozícióban csak egy változás történhet, ez azonos a pontmutációk számával – vagyis speciális körülmények között értelmezhető. Informatív lehet a változatosságot formáló folyamatokra a heterozigotizátsnál tárgyalt körülmények között (Wright-Fisher modell).

Genetikai változatosság jellemzése – DNS szekvencia

- **Szekvencia diverzitás** (átlagos páronkénti távolság)
 - gyakoriság mellett távolság információ is, pl. eltérő pozíciók száma ($\rightarrow p$ -távolság)
 - ha n db szekvencia (egyed), és d_{ij} az i és j szekvencia távolsága, az átlagos páronkénti távolság:

$$\pi^* = \frac{2}{n(n-1)} \sum_{i < j} d_{ij}$$

- d_{ij} nem csak a tapasztalt páronkénti eltérést jelentheti (genetikai távolságok becslése modellekkel)
- **Nukleotid diverzitás** (π): $\pi = \pi^*/L$, ha L a pozíciók száma
 - egy pozícióra vonatkoztatott érték
 - minden pozíciót egy lokusznak tekintve h_L és π ekvivalenciája
 - folyamat (Wright-Fisher modell):

$$E(\pi) = \theta$$

Két szekvenciát összehasonlítva az eltérést számos különböző módon számszerűsíthetjük. Legegyszerűbb esetben **távolságuk** az eltérő pozíció szám, ezt leosztva a szekvencia hosszával kapjuk a két szekvencia **p-távolságát**. Ez utóbbi így már egy szekvencia pozícióra vonatkoztatott érték. Emellett a **szekvencia evolúció modelljeivel** például a különböző eltérés típusok (pl. tranzíció és transzverzió) között is súlyozhatunk, vagy figyelembe vehetjük azt is, hogy egy pozícióban több változás történhet. Kiszámolva a távolságot az összes lehetséges szekvencia párra és ezt átlagolva kapjuk a szekvencia vagy egy nukleotidra kifejezve a nukleotid diverzitást. Ez utóbbi várható értéke ismét θ a Wright-Fisher modell feltételei mellett. Vegyük észre, hogy ezekkel a mérőszámokkal allél és haplotípus adatokból indultunk ki. Az ezeken alapuló következtetések gyakran nem igényelnek információt a genotípusról, vagy az egyedet haploidnak, homozigótának tekinthetjük ismert genetikai sajátosságokra alapozva (pl. mtDNS és rDNS szekvenciák). Továbbá a kettősszalú DNS csak egyik szálát nézzük, hiszen ennek ismeretében a másik (reverz komplementer) is ismert.

Genetikai változatosság jellemzése – DNS szekvencia

Példa:

Egyed	Haplotípus	Szekvencia
A	1	A G A T C
B	2	T G A T C
C	2	T G A T C
D	3	T G T T C
E	4	T C A C C
F	5	T C A C G

minta: 6 egyed, $m = 5$ allél (haplotípus), a szekvencia hossza $L = 5$ bp

allél gyakoriságok (p): $(1/6, 2/6, 1/6, 1/6, 1/6)$

haplotípus diverzitás: $h = 0.933$

szegregálódó pozíció szám: $S = 5$

átlagos páronkénti távolság: $\pi^* = 31/15 = 2.067$ (pl. $d_{12} = 1$)

nukleotid diverzitás: $\pi = 2.067/5 = 0.413$

A példa a különböző mérőszámok becslését szemlélteti szekvenciákra. Hasonlóan történik nagy mintára és hosszabb szekvenciákra is, erre különböző szoftverek állnak rendelkezésre.

Ellenőrző kérdések

- ❶ Mekkora az A allél relatív gyakorisága, ha a genotípus gyakoriságok:
 $f_{AA} = 0.3$, $f_{Aa} = 0.5$, $f_{aa} = 0.2$?
- ❷ Mit értünk mendeli karakter alatt?
- ❸ Mit értünk heterozigotitás alatt?
- ❹ Mekkora az átlagos heterozigotitás, ha három mikroszatellit lokuszra az alábbi heterozigotitás értékeket kaptuk: 0.980, 0.975, 0.985?
- ❺ Mit értünk szegregáló pozíció szám alatt?
- ❻ Mit értünk nukleotid diverzitás alatt?

**JELEN TANANYAG A SZEGEDI TUDOMÁNYEGYETEMEN
KÉSZÜLT AZ EURÓPAI UNIÓ TÁMOGATÁSÁVAL. PROJEKT
AZONOSÍTÓ: EFOP-3.4.3-16-2016-00014**

SZÉCHENYI 



MAGYARORSZÁG
KORMÁNYA

Európai Unió
Európai Szociális
Alap



BEFEKTETÉS A JÖVŐBE