


EFOP-3.4.3-16-2016-00014



---




**Pénzes Zsolt, Markó Bálint**

**Molekuláris ökológia: a koalíció**

Jelen tananyag a Szegei Tudományegyetemen  
készült az Európai Unió támogatásával.

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

Szegei Tudományegyetem  
Cím: 6720 Szege, 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 modern populációgenetika egy meghatározó koncepciójának, a koalíció elméletnek az alapjait tárgyaljuk kettes mintanagyságra, kitérve néhány alapvető genetikai változatosság jellemző értelmezésére is.

Ajánlott irodalom: Nielsen R, Slatkin M (2013) An Introduction to Population Genetics. Sinauer. (3. fejezet)

Az olvasólecke átnézése kb. 20 percet vesz igénybe.

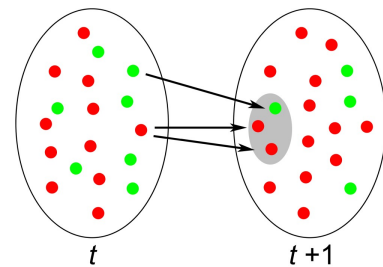
## Koaleszcencia

- Humán populációk, szekvencia adatok (pl. Wall et al 2008)
  - Európa 0.08%, Afrika 0.12% átlagos egyedek közötti különbség – következtetés a populációkra
  - pl. Wright-Fisher (WF) modell – allél gyakoriság változás
  - de: alkalmazása körülményes. . .
- Koaleszcencia elmélet (1980-as évektől)
  - modelleken alapul, pl. Wright-Fisher modell
  - logika: minta genealógia – jelenből időben visszafelé, leszármazási sorok „összeolvadása” (coalescence)
  - tradicionálissal azonos eredmények - de számos előny
  - adatok és elmélet közvetlen kapcsolata
  - általában szélesebb körű alkalmazhatóság
- Cél: az elmélet logikai alapjai

Egy populáció **neutrális genetikai változatosságát** a mutációk és genetikai sodródás elmélete alapján értelmeztük, azzal a céllal, hogy a tapasztalt változatosságból a mintázott populációk sajátosságaira tudjunk következtetni. Például intergénikus DNS szekvenciákat elemezve azt találták, hogy az afrikai humán populációk (biaka, san) genetikai változatossága nagyobb, mint az európai (baszk) vagy ázsiai (han kínai) populációké, ami a *Homo sapiens* afrikai eredetével konzisztens eredmény. Az elmélet ismeretében különböző hipotéziseket (modelleket) formalizálhatunk egy jelenség magyarázatára. A hipotézisek között az alapján döntünk, hogy milyen mértékben magyarázzák a tapasztalt mintázatot. Azonban a tradicionális modellek – allél gyakoriság időbeli változásának modellezése egy ősi populációból kiindulva – illesztése az adatokra gyakran körülményes. Molekuláris (főleg szekvencia) adatok elemzésében a **koaleszcencia** elmélet jelentős előrelépést hozott. A hangsúlyt a minta genealógiájára fektetjük, a mintavétel időpontjából indulunk ki. Ezzel a megközelítési móddal kézenfekvőbb egyes kulcsfontosságú sajátosságok jelentőségének megítélése, és a hagyományos megközelítési mód eredményeit is sokszor egyszerűbben tudjuk értelmezni, amellett hogy a modellek és eredmények is gyakran szélesebb körben alkalmazhatóak.

## Koaleszcencia, $n = 2$

- Ősök „követése” a populációban – haploid WF modell
  - 2 „egyed” (allél) kiválasztása a  $t + 1$  generációban (minta,  $n = 2$ ), a populáció mérete  $2N$  (konstans)
  - azonos várható utódszám  $\rightarrow$  szülő bármelyik előző generációbeli ( $t$ ) egyed lehet azonos valószínűséggel:  $1/(2N)$
  - két egyed azonos szülőtől – valószínűsége:  $1/(2N)$ 
    - két leszármazási sor „összeolvadása” a  $t$ . generációban
    - vagyis két allél másolat egy ősi allélról (azonos eredet, IBD)
  - vagy különböző szülőktől – valószínűsége:  $1 - 1/(2N)$ 
    - de közös ős a  $t - 1$  ( $t - 2, t - 3$  stb.) generációban – valószínűsége?
    - igazolható, hogy két leszármazási sornak mindig lesz közös őse
- WF modell
- minta ( $t + 1$ . generációban): 3 egyed
- két egyed ua. szülőtől (koaleszcencia a  $t$ . generációban)



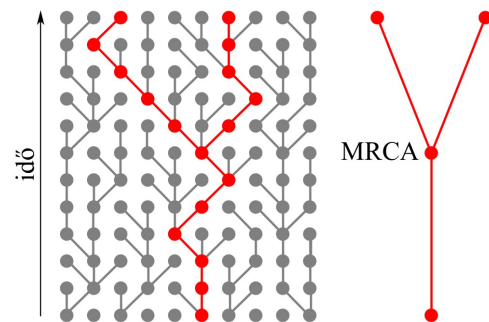
A Wright-Fisher populáció bármely „egyedének” (ténylegesen egy allél, egy szekvencia) az őse azonos eséllyel lehet bármelyik előző generációbeli egyed. Vagyis egy kiválasztott egyedre vonatkozóan (minta) az előző generáció bármely egyedére ennek valószínűsége  $1/(2N)$ , ha a populáció mérete  $2N$ .

Két egyedet kiválasztva (minta, melynek nagysága  $n = 2$ ) mindkettő származhat ugyan attól a szülőtől. Ha ez történik a két leszármazási sor 1 generáció alatt összeolvad egy közös ősből. Nem lényeges most számunkra az, hogy melyik egyed a közös ős, kérdésünk pusztán arra vonatkozik, hogy egy szülőtől származnak-e. Egy a  $t + 1$  generációból vett mintára (jelen), annak a valószínűsége hogy közös ősük  $t$  generációbeli (azonnali koaleszcencia)  $1/(2N)$  (hiszen egy adott közös szülő valószínűsége  $1/(2N)^2$ , de ez a  $2N$   $t$  generációbeli egyed bármelyike lehet). Különösen nagy populációban ez valószínűtlen ( $1/(2N)$  kicsi, pl.  $2N = 500$  esetén 0,2%). Annak a valószínűsége, hogy közös ősük nem előző ( $t$ ) generációbeli:  $1 - 1/(2N)$ .

Ez utóbbi esetben a közös őst egy korábbi ( $t - 1, t - 2$  stb.) generációban keressük. Igazolható, hogy a két leszármazási sornak **mindig van közös őse** – kérdésünk az, hogy hány generációval ezelőtti.

## Koaleszcencia, $n = 2$

- MRCA (Most Recent Common Ancestor): az utolsó közös ős – mindig létezik...
- A történet szemléltetése: koaleszcens fa
  - ág (él): leszármazási sor – hossza: idő
  - levél (terminális csomópont): gén másolat (minta)
  - belső csomópont: közös ős – MRCA a fa gyökere
- Koaleszcencia idő: a közös őstől eltelt idő – a keresett jellemző
- populáció, 10 db haploid „egyed”
- egymást követő generációk (12) és leszármazási kapcsolatok
- 2 elemű **minta** (jelen) és a **koaleszcencia fa**
- ábra:  $t_{MRCA} = 5$



A mintánk legyen a jelen populációból származó **két szekvencia**. Elsősorban ezt az esetet tárgyaljuk a továbbiakban, az erre kapott eredményekből tudunk majd következtetni a nagyobb minta nagyságra is.

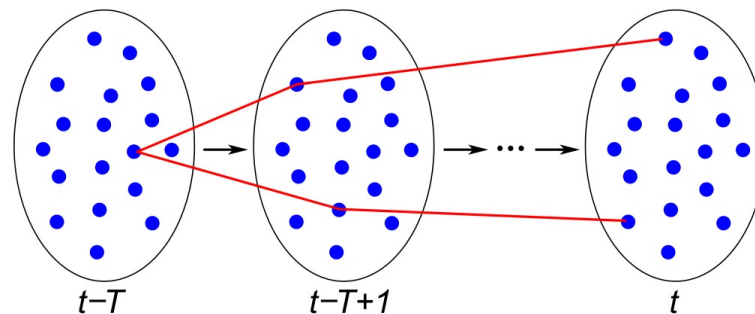
A koaleszcencia események sorozatát **koaleszcens fával** ábrázoljuk, melynek terminálisai (levelek) a mintánk elemei, belső csomópontjai a közös ősök, amelyben a leszármazási sorok összeolvadnak. Mintánk minden elemének **utolsó közös őse** a fa gyökere (MRCA), a minta minden eleme ebből származtatható. Természetesen ez utóbbi ősei is közös ősök, azonban az utolsó kiemelt jelentőségű lesz. A koaleszcens fa ilyen módon genealógiát, leszármazási kapcsolatokat szemléltet.

A jelenből kiindulva a koaleszcenciáig eltelt időt **koaleszcencia időnek** nevezzük, amelyet az ábrán generáció számban adtuk meg. Öt generáció alatt következett be a koaleszcencia ( $t_{MRCA} = 5$ ) ebben az esetben. Egy másik populációban vagy egy másik mintára a populációból ez bekövetkezhetett volna azonnal vagy akár később (a jelenből kiindulva), melynek a valószínűsége adott feltételek mellett számolható – a **koaleszcencia idő eloszlása** meghatározható.

## Koaleszcencia idő, $n = 2$

- $t_{MRCA}$  becslése – hogyan? ( $n = 2$  eset, diszkrét generációk)
  - 1 azonos szülő (két allél közös őse, koaleszcencia) valószínűsége az előző generációban:  $1/(2N)$
  - 2 eltérő szülő (nincs koaleszcencia az előző generációban):  $1 - 1/(2N)$
  - 3 további generációk hasonlóan – pl.  $r$  generáción keresztül nincs közös őse (függetlenség):  $[1 - 1/(2N)]^r$
  - 4 MRCA valószínűsége az  $r$ . generációban (és  $r - 1$  generáción keresztül nincs közös szülő, geometriai eloszlás):  $[1 - 1/(2N)]^{r-1} [1/(2N)]$

WF mintázás időben visszafelé ( $T$  koaleszcencia idő,  $n = 2$ ):



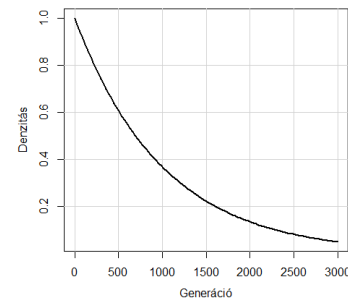
A **koaleszcencia idő** diszkrét Wright-Fisher modellel **geometriai eloszlású**, amely az alábbi módon származtatható. A mintánk a  $t$  generációból származik.

1. Annak a valószínűsége, hogy a két elemű minta közös őse előző generációbeli  $1/(2N)$ , ahogy már korábban beláttuk. Ez esetben  $t_{MRCA} = 1$ .
2. Ha nem volt azonnali koaleszcencia (valószínűsége  $1 - 1/(2N)$ ), az azt megelőző generációban is bekövetkezhet ettől *függetlenül*, azonos valószínűséggel ( $1/(2N)$ ,  $2N$  konstans).
3. Vagy bármely generációban, feltéve hogy korábban nem következett be (hiszen kettes minta nagyságra csak egyszer fordulhat elő). Annak a valószínűsége, hogy  $r$  generáción keresztül nincs közös őse:  $[1 - 1/(2N)]^r$
4. Annak a valószínűsége, hogy a koaleszcencia az  $r$ . generációban következik be (és  $r - 1$  generáción keresztül nincs közös őse):  $[1 - 1/(2N)]^{r-1} [1/(2N)]$

Ez alapján bármely generációra számolható a koaleszcencia valószínűsége. A koaleszcencia idő várható értéke  $2N$ . De mielőtt ezt tovább értelmeznénk, érdemes bevezetni egy további egyszerűsítést.

## Koaleszcencia idő, $n = 2$

- Nagy populáció ( $N \rightarrow \infty$ ) – matematikai egyszerűsítés...
  - általános eredmények
  - időben folytonos, időt  $2N$  generációban mérjük:  $t = r/(2N)$
  - nincs közös ő 2Nt generáción keresztül:  $e^{-t}$
  - a koaleszcencia idő exponenciális eloszlású
    - paramétere (ráta,  $\lambda$ ) és várható értéke ( $1/\lambda$ ):  $\lambda = 1/\lambda = 1$
    - a koaleszcencia idő várható értéke  $E[t] = 1$  az időt  $2N$  generációban mérve („átlagos várakozási idő”)
    - a koaleszcencia idő várható értéke  $2N$  generáció és konstans  $1/(2N)$  koaleszcencia ráta (koaleszcencia valószínűsége/generáció)
- De realizációk (populáció) – jelentős különbségek (variancia)
  - generációk számának eloszlása a koaleszcenciáig ( $2N = 1000$ )
  - recens koaleszcencia valószínűsége nagyobb
  - koaleszcencia bekövetkezési esélye csak az időintervallum hosszától függ (feltéve hogy korábban nem következett be)



Feltételezve hogy a populáció mérete nagy ( $N \rightarrow \infty$ ), egy további egyszerűsítéssel élhetünk. Az időt folytonosnak tekintjük és  $2N$  generációban mérjük (bevezetve a  $r = 2Nt$  jelölést), így a koaleszcencia idő eloszlását 1 paraméterű **exponenciális eloszlással közelítjük** (formálisabban:  $r = 2Nt$  idő alatt nincs koaleszcencia  $[1 - 1/(2N)]^r$  valószínűséggel, a fenti feltétellel a  $[1 - 1/(2N)]^{2Nt}$  sorozat határértéke  $e^{-t}$ ). Kettes mintanagyságra a koaleszcencia idő várható értéke  $2N$  generáció, vagy az időt generációban mérve  $E[t] = 1$ . Az átskálázott időléptékkel lényegesen egyszerűbb lesz a különböző jellemzőket értelmezni. Noha a várható koaleszcencia idő  $2N$  generáció, azonban ettől még egy populációban ennél sokkal hamarabb vagy később is bekövetkezhet, jelentősek lehetnek a különbségek (szórás is  $2N$  generáció). Annak a valószínűsége, hogy a koaleszcencia idő  $t$ -nél kisebb ( $0 - t$  intervallumban):  $1 - e^{-t}$ . A közelítés alkalmazása nem jelenti azt, hogy pl. kis populációkra nem használható a modell.

A **standard koaleszcencia folyamattal** a leszármazási folyamatot modellezzük időben visszafelé haladva. Az eredmények a tradicionális Wright-Fisher modellnél általánosabbak, például nagy véletlenszerűen kombinálódó diploid populációra is hasonló következtetésekre juthatunk.

## Mutáció, $n = 2$ eset

- $\mu$  mutációs valószínűség generációnként  
 →  $\mu r$  mutációt várunk  $r$  generáció alatt  
 →  $2N\mu t$  mutáció egy  $t$  hosszú leszármazási sorban
- Várható mutációk száma két leszármazási sorban ( $n = 2$ ):  $2 * 2N\mu t$
- Két allélt elkülönítő várható mutáció szám (közös őstől,  $E[t]=1$ ):  
 $\theta = 4N\mu$ 
  - $\theta$  kiemelt szerepű populáció paraméter, várható mutáció szám
  - populáció mérete és a genetikai változatosság kapcsolata
  - kis populáció → átlagosan kis koaleszcencia idő és kis változatosság (kevesebb mutáció halmozódott fel)
- Tetszőleges  $\tau$  leszármazási sor hossza a várható mutáció szám  
 $2N\mu\tau = \tau\theta/2$ 
  - elméleti jelentősége: koaleszcencia és mutáció folyamata (a legegyszerűbb esetben) szétválasztható
  - $\theta/2$  generációnkénti mutációs ráta a koaleszcens fán
  - szelekcióval már összetettebb

Tételezzük fel, hogy **új mutáció** generációnként  $\mu$  valószínűséggel következhet be (mutációs ráta) bármely leszármazási sorban. Ez egy  $t$  hosszú leszármazási sorban (az időt  $2N$  generációban mérve,  $r = 2Nt$ )  $2N\mu t$  új mutációt jelent. Ha  $t$  a koaleszcencia idő, együttesen a két leszármazási sorban szétválásuk óta (vagy koaleszcenciájukig a jelenből nézve) összesen  $4N\mu t$  mutáció halmozódik fel. Mivel  $E[t] = 1$ , a várható mutáció szám, amely a mintánk két szekvenciáját elkülöníti  $\theta = 4N\mu$ .

$\theta$  a populáció genetikai változatosságának egy alapvető mérőszáma. A genetikai változatosság és a populáció méret között egyszerű kapcsolatot definiál. Kis populációban a várható koaleszcencia idő kisebb, így kevesebb mutáció halmozódik fel, ezért a változatosság is kisebb egy nagy populációhoz képest.

Legyegegyeszerűbb esetben a mutáció és koaleszcencia folyamat szétválasztható, amely jelentős elméleti és technikai egyszerűsítést jelent. Egy  $\tau$  hosszú leszármazási sorra (az időt  $2N$  generációban mérve), a koaleszcens fa  $\tau$  hosszú szegmensére, a mutációk várható száma  $\tau\theta/2$ . Érdemes azonban megjegyezni, hogy a koaleszcencia kiterjesztése természetes szelekció irányába (vagyis nincs neutralitás) általában már nem tehető meg ilyen egyszerűen.

## $\theta$ becslése

- Cél: változatosság adatok és modellek összekapcsolása  $\rightarrow \theta$  becslése
- DNS szekvenciák – a végtelen pozíció modell
  - minden mutáció új változatosságot mutató helyet (SNP) jelent (szekvencia hossza végtelen) – gyakorlatban ritkán...
  - szegregálódó helyek (ábra: 10, 14, 21 pozíciók)  $\rightarrow$  mutációk száma meghatározható (ábra: 3 pozíció, 3 különböző haplotípus)
  - nukleotidok között nem tettünk különbséget (bináris adatmátrix)
- Páronkénti különbségek átlagos száma ( $\pi$ ), ha  $d_{ij}$ : a nukleotid eltérések száma  $i$  és  $j$  szekvencia között (példa:  $n = 5$ ,  $\pi = 1,6$ )

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

1. szekv.	GGATCTGATAGAAACCTGGAAAGACACACAGATAGACAAG	000
2. szekv.	GGATCTGATAGAAACCTGGATAGACACACAGATAGACAAG	011
3. szekv.	GGATCTGATAGAAACCTGGATAGACACACAGATAGACAAG	011
4. szekv.	GGATCTGATGGAACCTGGATAGACACACAGATAGACAAG	101
5. szekv.	GGATCTGATGGAACCTGGATAGACACACAGATAGACAAG	101

Ahhoz, hogy a modellünket az adatokkal is össze tudjuk kapcsolni, a **mutációk specifikusabb modelljére** van szükség. Szekvencia adatokra induljunk ki a végtelen pozíció modellből, ahol minden pozíció maximum csak egyszer módosulhat. Ez azt jelenti, hogy megközelítési módunk a gyakorlatban extrém kis mutációs ráta esetén használható, kis időléptékben, ahol a többszörös mutációt elhanyagolhatjuk (ennek figyelembe vételére persze léteznek modellek, lásd filogenetika). A végtelen pozíció modellben az eltérő pozíció szám (a tapasztalt mintázat) egyben a bekövetkezett mutáció számot mutatja.

A változatosság mérőszámai kapcsán már megismertedtünk a **páronkénti különbségek átlagos számával** (szekvencia diverzitás,  $\pi$ ), amely a változatosság egyik leggyakrabban használt jellemzője szekvencia adatokból (általában egy pozícióra vonatkoztatott formában, lásd nukleotid diverzitás). Legyen  $d_{ij}$  az eltérő pozíciók száma a minta két ( $i$  és  $j$ ) szekvenciája között.  $n$  db szekvencia esetén  $\binom{n}{2} = n(n-1)/2$  szekvencia pár képezhető,  $\pi$  ezek különbségeinek az átlaga. A példában  $n = 5$ , így 10 pár képezhető, és  $\pi = (2 + 2 + 2 + 2 + 0 + 2 + 2 + 2 + 2 + 0)/10 = 1,6$  ( $d_{12} = 2$ ,  $d_{13} = 2$  stb.).



## $\theta$ becslése

- Végtelen pozíció modell:  $d_{ij}$  = mutációk száma a koaleszcencia előtt  
→ várható értéke:

$$E[d_{ij}] = \theta$$

- $\theta$  becslése egy tetszőleges szekvencia pár közötti átlagos különbség alapján...
- Becslés az összes szekvencia pár alapján, párok átlaga ( $E[\pi]$ ) – több információ

$$E[\pi] = \theta$$

- $\theta$  torzítatlan becslése szekvencia adatokból
- Becslés  $\pi$  alapján – a Tajima becslő (Tajima's estimator),  $\hat{\theta}_T$

Két szekvencia közötti különbség a közös őstől történt szétválás óta (vagy jelenből kiindulva a koaleszcencia eseményig) bekövetkezett mutációk számának felel meg a végtelen pozíció modellben. Vagyis a modell alapján akár egy szekvenciapárból is becsülhető a  $\theta$  paraméter értéke (ami korábbi definíció szerint a várható felhalmozódott mutáció számot jelenti):  $E[d_{ij}] = \theta$ . Azonban több szekvencia pár alapján, ezek átlagával jobb becslés adható:

$$E[\pi] = E \left[ \frac{\sum_{i < j} d_{ij}}{n(n-1)/2} \right] = \frac{\sum_{i < j} E[d_{ij}]}{n(n-1)/2} = \frac{\theta n(n-1)/2}{n(n-1)/2} = \theta$$

$\theta$  átlagos páronkénti különbségekkel történő becslőjét Tajima-becslőnek nevezzük (jele  $\hat{\theta}_T$ ).  $\hat{\theta}$  a mintából számolt statisztika (becslő) amivel a populációnk genetikai változatosságát jellemző  $\theta$  paraméter becsüljük. Levezetésünk alapján a végtelen pozíció modell feltételei mellett ( $d$  a tapasztalt mutáció szám), kis tapasztalt változatosságra várunk megbízható becslést, azonban ennél összetettebb esetekre is kiterjeszhető. Szekvencia adatokból a  $\theta$  paraméter egyéb módon is becsülhető.

## Heterozigozitás, $n = 2$ eset

- A végtelen allél modell – cél: heterozigozitás értelmezése
  - minden mutáció új allélt eredményez (allél szám végtelen), vagyis IBS (állapot, megfigyelt)  $\equiv$  IBD (eredet, nincs mutáció)
  - csak allél típus (haplotípus) kategóriák (nominális) – azonos vagy eltér információ két allélra
  - pl. enzimpolimorfizmus, „típusok” kapcsolata ismeretlen (vö. szekvenciák távolsága), SNP
- Két azonos allél mintázásának valószínűsége (= nincs mutáció, végtelen allél modell) a koaleszcencia folyamattal?
  - koaleszcencia és mutáció – független exponenciális vv.-k
  - mutációs ráta mindkét leszármazási sorban:  $\theta/2$
  - koaleszcencia ráta és koaleszcencia idő várható értéke: 1
  - koaleszcencia következik be előbb (jelenből nézve) – valószínűsége:

$$\frac{1}{1 + 2\theta/2} = \frac{1}{1 + \theta}$$

Mutációk egy másik gyakran alkalmazott modellje a **végtelen allél modell**, ahol minden mutáció új allélt eredményez. Akkor használjuk (például enzimpolimorfizmus adatokra, de gyakran önálló SNP-kre is), ha mindössze típusokat (állapotokat) különítünk el egymástól. A modellben tehát egy új állapot megjelenése egy mutációt, az azonos állapot (IBS, identity by state) azonos eredetet (IBD, identity by descent) jelent.

A koaleszcencia folyamat alapján annak a valószínűségét keressük, hogy mintánk két eleme azonos típusú marad, vagyis a koaleszcenciáig (jelenből kiindulva) nincs mutáció egyik leszármazási sorban sem a végtelen allél modellnek megfelelően. Az exponenciális eloszlású koaleszcencia idő várható értéke és a generációnkénti koaleszcencia ráta egyaránt egy, míg a mutációs ráta  $\theta/2$  generációnként mindkét leszármazási sorra (az időt  $2N$  generációban mérve). A mutációt és koaleszcenciát független random folyamatoknak tekintve annak a valószínűsége, hogy a koaleszcencia következik be előbb (a jelenből kiindulva) számolható, ez a koaleszcencia ráta osztva a teljes (koaleszcencia + mutáció mindkét leszármazási sorra) rátával.

## Heterozigotitás, $n = 2$ eset

- A két allél azonos  $\rightarrow$  a várható homozigotitás (véletlenszerű kombinálódással egy lokuszon)
- Várható heterozigotitás:

$$E[h] = 1 - \frac{1}{1 + \theta} = \frac{\theta}{1 + \theta}$$

- random minta egy véletlenszerűen kombinálódó populációból – két eltérő allél valószínűsége
- egy egyedben a várható lokusz arány, amelyre heterozigóta
- ha a standard koaleszcencia modell és a végtelen allél modell igaz. . .
- Hasonló eredmények a végtelen pozíció modellel is
  - haplotípus heterozigotitás = két eltérő haplotípus valószínűsége. . .
- De  $IBS \neq IBD$ 
  - pl. back-mutáció: azonos állapot (IBS), de mutációk  $\rightarrow$  nem azonos eredet (IBD)
  - több mutáció, független eredet  $\rightarrow$  szekvencia evolúció modelljei

Ha a koaleszcencia következik be előbb, a két allél azonos a kérdéses lokuszon. Ez utóbbi valószínűsége nem más, mint egy véletlenszerűen kombinálódó populáció **várható homozigotitása**. Mivel a várható homozigotitás és heterozigotitás összege 1, a **várható heterozigotitás**  $\theta/(1 + \theta)$ , a genetikai változatosság egy általánosan használt jellemzője. A várható heterozigotitás jelentése, amennyiben a standard koaleszcencia és a végtelen allél modell igaz:

- véletlenszerű kombinálódással (párválasztással) a heterozigóta lokuszok aránya
- . . . és annak a valószínűsége, hogy egy lokuszon a két allél különbözik
- vagy a populációban egy lokuszra vonatkozóan két véletlenszerűen kiválasztott allél különbözik (**haplotípus heterozigotitás**)

Természetesen egy mutáció nem szükségszerűen eredményez új allélt a gyakorlatban. A hatása alapján elkülönített allél megjelenése mögött különböző mutációk állhatnak, sőt az azonos állapot is lehet pl. back-mutáció következménye. Ennek figyelembe vételére különböző szekvencia evolúciós modellek léteznek, ahogy a végtelen pozíció modellel kapcsolatban is említésre került.

## Ellenőrző kérdések

- 1 Mit értünk koaleszcencia alatt?
- 2 Mi a koaleszcencia idő?
- 3 Milyen eloszlású a koaleszcencia idő?
- 4 Mekkora a generációnkénti koaleszcencia ráta? ( $n = 2$  eset)
- 5 Mekkora a várható koaleszcencia idő 2 szekvenciára?
- 6 Mi a végtelen pozíció modell?
- 7 Mi a végtelen allél modell?
- 8 Mi a  $\theta$ , hogyan értelmezzük a koaleszcencia alapján?
- 9 Mit értünk MRCA alatt?
- 10 Mi a várható heterozigotizáció?
- 11 Egy populáció palacknyakon ment keresztül 100 generációval ezelőtt, majd ezt követően a populáció mérete 500 egyed és Wright-Fisher populációként evolválódott tovább. Mekkora két allél palacknyak előtti koaleszcenciájának a valószínűsége (jelenből kiindulva)?



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](http://www.u-szeged.hu)  
[www.szechenyi2020.hu](http://www.szechenyi2020.hu)



MAGYARORSZÁG  
KORMÁNYA

Európai Unió  
Európai Szociális  
Alap



BEFEKTETÉS A JÖVŐBE