Week 8 Population Genomics

Friday, 11 March 2022   8:05 AM

HWE   1 genomic locus; diploid;

bi-allelic $\{A, G\}$

GENOTYPES: AA, AG, GG

n = # of samples we have sequenced

$\Rightarrow$ $2n$ = total # of alleles

in gen. t; $P^t(AA) = \dfrac{n_{AA}}{n}$; $P^t(AG) = \dfrac{n_{AG}}{n}$

$P^t(GG) = \dfrac{n_{GG}}{n}$

$\left. \begin{array}{l} P = P^t(A) = P^t(AA) + \dfrac{1}{2} P^t(AG) \\[3mm] q = P^t(G) = P^t(GG) + \dfrac{1}{2} P^t(AG) \end{array} \right]$ $\begin{array}{c} P^t(A) + P^t(G) \\ = 1 \\ \left[P + q = 1 \right] \end{array}$

$\hookrightarrow$ randomly mating $\hookrightarrow$ no selection

$\hookrightarrow$ no new mutations ✗ $\hookrightarrow$ no migration

$\hookrightarrow$ ∞ ly large population ✗

$\Rightarrow$ I] $P^{t+1}(AA) = P^2$; $P^{t+1}(AG) = 2pq$; $P^{t+1}(GG)$

$= q^2$

II] $P^{t+1}(A) = P^t(A) = P$

$P^{t+1}(G) = P^t(G) = q$

---

i] FINITE POPULATION ; n ⌣   DRIFT

$\overset{\smile)}$

$$p^t(A) = P$$
$$p^t(G) = 1-P$$

$\longrightarrow$

$$p^{t+1}(A) = \frac{x}{2n}$$
$$p^{t+1}(G) = \frac{2n-x}{2n}$$

generation t

gen. $t+1$

$t+1$

$P(2n$ total alleles; x of A, $2n-x$ of $G)$

$$= \ ^{2n}\!\left(\begin{matrix} \\ x \end{matrix}\right. \underset{x}{P}^x \ (1-P)^{2n-x} \quad \Big]$$

∞ ly large populations $\rightarrow$ no drift

finite populations $\rightarrow$ lots of
(small) drift

MUTATIONS

μ = mutation rate of $A \leftrightarrow G$

$$P(A \leftrightarrow G) = \mu \qquad [\text{bidirectional}]$$

$$E[\ p^{t+1}(A)\ ] \ = \ \overset{t}{P}\overset{P}{(A)} \ (1-\mu) \ + \ p^t(G)\,\mu \quad\Big]$$
$$\underset{P}{\uparrow\ \uparrow} \qquad\qquad\qquad\qquad \underset{1-P}{\nearrow} \ \ L①$$

$= $ Prob. of observing
an A allele & it didn't
mutate to G

$+ (OR)$ Prob. of observing a G
allele & it did mutate
to A

$$E[\ p^{t+1}(G)\ ] \ = \ 1 - E[\ p^{t+1}(A)\ ]$$

$$= \ p^t(G)\,(1-\mu) \ + \ p^t(A)\,\mu$$

MUTATION - DRIFT EQUILIBRIUM

$$P^{t+1}(A) = P^t(A) = P$$

Plugging this into ①

$$P = P(1-\mu) + (1-P)\mu$$

$$\cancel{P} = \cancel{P} - P\mu + \mu - \mu P.$$

$$2\mu P = \mu$$

$$P = \frac{\mu}{2\mu} = \frac{1}{2}$$

MUTATION PROB.

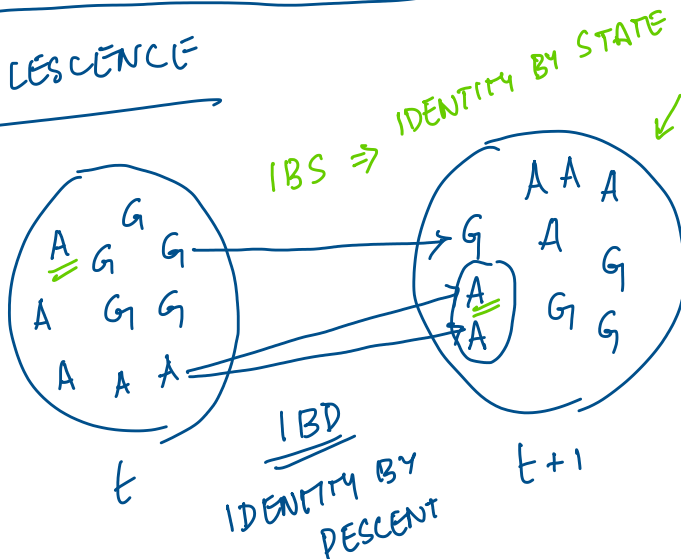$$\mu_1 = P(A \to G)$$

$$\mu_2 = P(G \to A)$$

$$P' = \frac{\mu_2}{\mu_1 + \mu_2}$$

generalized
mutation
- drift

COALESCENCE

IBS ⇒ IDENTITY BY STATE

IBD
IDENTITY BY
DESCENT

t   t+1

2n
total
alleles in
each generation

$[n = N]$

... ...pies are IBD in

$$P(2 \text{ gene not} \atop \text{prev. generation}) = \frac{1}{2N}$$

$\Rightarrow P(2 \text{ gene copies are not IBD} \atop \text{in prev. generation}) = 1 - \frac{1}{2N}$

if I have 'n' generations

$$P(2 \text{ gene copies don't coalesce in} \atop r \text{ generations}) = \left(1 - \frac{1}{2N}\right)^n$$

$P(2 \text{ gene copies don't coalesce in} \atop n \text{ generations } \& \text{ then they} \atop \text{coalesce}) = \underbrace{\left(1 - \frac{1}{2N}\right)^r} \underline{\left(\frac{1}{2N}\right)}$
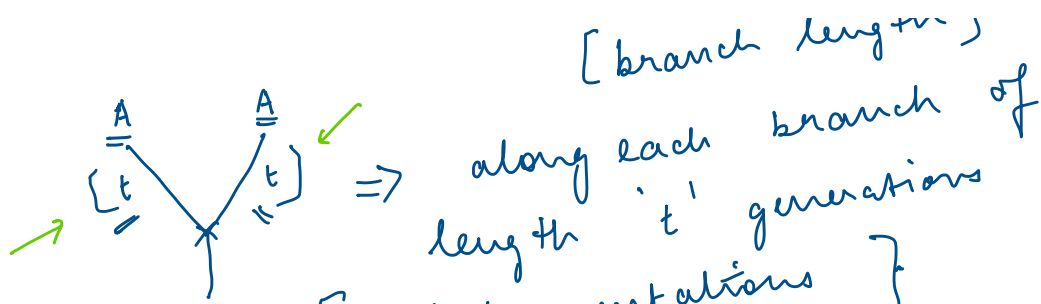
$$\boxed{n = 2Nt}$$

as $N \to \infty$

$$\left(1 - \frac{1}{2N}\right)^{2Nt} \longrightarrow e^{-t}$$

$\Rightarrow \left[\text{expected coalescence time} = 2N \atop \text{generations} \right]$

$\mu = $ mutation rate / generation.

$$n = 2Nt \quad \Rightarrow \quad t = \frac{n}{2N}$$

$\Rightarrow$ along each branch of length 't' generations [branch lengths]

[$2N\mu t$ mutations]

$\Rightarrow$ b/w any 2 lineages, time $= 2t$

$\Rightarrow$ total # of mutations accumulated b/w any 2 lineages $= 2 \times 2N\mu t$

$= 4N\mu t$

$$\boxed{\theta = 4N\mu}$$

population size scaled mutation rate

$\hookrightarrow$ genetic diversity $\leftarrow$

for any branch of length 'τ'

expected # of mutations $= 2N\mu\tau$

$= \dfrac{\tau\theta}{2}$

— (2)

INFINITE SITES MODEL

$\rightarrow$ each mutation creates a new variable site (SNP)

variable sites at a locus in a population

$\begin{array}{l} 1 \quad T \; A \; G \; A \; A \; C \;] \\ 2 \quad T \; A \; G \; A \; A \; A \end{array}$

$\hookrightarrow$ SEGREGATING SITES

3 T A G A A A
4 T G G A A C
5 T G G A A C

$S = 2$

TAJIMA ESTIMATOR (π)

$=$ average # of pairwise differences in 'n' sequences [eg. $n = 5$)

total # of pairwise comparisons $= \dfrac{n(n-1)}{2}$

$$\pi = \frac{\sum\limits_{i<j} d_{ij}}{n(n-1)/2} = \hat{\theta}_T$$

eg. $$\pi = \frac{1 + 0 + 2 + 0 + 1 + 2 + 2 + 1 + 2 + 1}{5 \times 4/2}$$

$$= \frac{12}{10} = 1.2 = \hat{\theta}_r$$

EFFECTIVE POPULATION SIZE $= N_e$

is the size of a WF population that is evolving at the same rate at the natural population