

Lecture 13. Synonymous and non-synonymous rates. Two tests of neutrality. Recombination and polymorphism.

4.7 Synonymous and non-synonymous rates

Genetic code is redundant Table 8.2, p. 339

three types of sites: 0 = non-degenerate site

2 = two-fold site and 4 = four-fold site

At a two-fold site $\frac{1}{3}$ of substitutions are synonymous

Effective numbers of sites

$$l_s = l_4 + \frac{1}{3} \cdot l_2 \text{ and } l_n = l_0 + \frac{2}{3} \cdot l_2$$

$$\text{total length } l = l_0 + l_2 + l_4 = l_s + l_n$$

Fig 8.12, p. 341 and Fig 8.13, p. 342

different substitution rates $\lambda_s = \frac{d_s}{2t}$ and $\lambda_n = \frac{d_n}{2t}$

$$d_s = \frac{\text{synonymous changes}}{l_s} \text{ and } d_n = \frac{\text{nonsynonymous changes}}{l_n}$$

Usually $\lambda_s > \lambda_n$ because of deleterious mutations

Fig 8.14, p. 343: mammalian nuclear DNA rates

Neutrality: $\lambda_s = \lambda_n$, positive selection: $\lambda_s < \lambda_n$

Genome averages

Wide variety of unconstrained substitution rates λ_s

plant chloroplast DNA	$1 \cdot 10^{-9}$
mammalian nuclear DNA	$3.5 \cdot 10^{-9}$
plant nuclear DNA	$5 \cdot 10^{-9}$
E.coli and Salmonella enterica bacteria	$5 \cdot 10^{-9}$
Drosophila nuclear DNA	$1.5 \cdot 10^{-8}$
mammalian mitochondrial DNA	$5.7 \cdot 10^{-8}$
HIV-1	$6.6 \cdot 10^{-3}$
Influenza A virus	$1.3 \cdot 10^{-2}$

Ex 2: bacterial gene

Observed differences per site

$$d_s = \frac{7}{10+12/3} = 0.5, \quad d_n = \frac{2}{38+12 \cdot 2/3} = 0.04$$

$$\text{uncorrected estimates } \tilde{\lambda}_s = 3.1 \cdot 10^{-9}, \quad \tilde{\lambda}_n = 0.3 \cdot 10^{-9}$$

Positive selection evidence

in a study of 3595 groups of homologous sequences only
 17 groups with λ_n/λ_s significantly larger than 1
 many of these are sex-related genes (favor speciation)

In some immunoglobulin genes

$\lambda_n/\lambda_s > 1$ in certain regions
 overdominant selection for antibody diversity

4.8 Two tests of neutrality

H_0 : observed polymorphism is due to selective neutrality
 of mutations and not due to natural selection

McDonald-Kreitman test

Chi-square test of homogeneity comparing two pairs of
 numbers of (synonymous, non-synonymous) differences

1. fixed differences between species
2. within species polymorphic sites

Reject the null hypothesis of neutrality

if two distributions are significantly different

Ex 12: G6PD gene in *Drosophila*

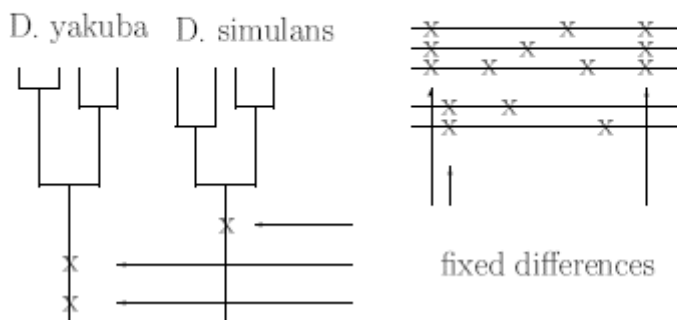
12 alleles in *D.yakuba* and 6 alleles in *D.simulans*

differences	between species	within species	total
synonymous	17(20.3)	29(25.7)	46
nonsynonymous	6(2.7)	0(3.3)	6
total	23	29	52

Excess of nonsynonymous fixed differences

positive selection of advantageous nonsyn. mutations

$\chi^2 = 8.6$, $df = 1$, $\sqrt{8.6} = 2.93$, $P = 0.0034$, reject H_0



Tajima test

tests neutrality using polymorphisms in one species
compares two estimates of θ : $\hat{\theta} = S/a_1$ and π
 S and π react differently on presence of selection
 S examines the number of polymorphic sites
 π assesses the site frequencies p_A, p_G, p_T, p_C

Very unequal p_A, p_G, p_T, p_C imply smaller π
almost equal p_A, p_G, p_T, p_C imply larger π

Ex 13: configuration and nucleotide diversity

$n = 12, l = 1$, number of pairs of sequences $\binom{12}{2} = 66$

config	(10,1,1,0)	(9,1,1,1)	(6,3,2,1)	(4,3,3,2)	(3,3,3,3)
#(pmm)	21	30	48	53	54
π	0.318	0.455	0.727	0.803	0.818

Tajima test statistic

Under hypothesis of neutrality $\text{Var}(\pi - \frac{S}{a_1}) = \frac{c_1\theta}{l} + c_2\theta^2$
where $c_1 = b_1 - \frac{1}{a_1}, c_2 = b_2 - \frac{n+2}{a_1n} + \frac{a_2}{a_1^2}$

$$D = \frac{\pi - S/a_1}{\sqrt{e_1 S + e_2 S(S-1/l)}}, \text{ where } e_1 = \frac{c_1}{a_1}, e_2 = \frac{c_2}{a_1^2 + a_2}$$

Null distribution of Tajima's D is tabulated by simulation
might be approximated by a Beta distribution
Significant $D > 0$ means almost equal p_A, p_G, p_T, p_C :
either balancing selection (overdominance) or
diversifying selection when rare alleles are favored
Significant $D < 0$ means very unequal p_A, p_G, p_T, p_C :
either selection against rare alleles or
recent bottleneck implying reduced genetic variation

4.9 Recombination and polymorphism

Fig 5.9, p. 182: evolutionary benefit of recombination

Fig 5.10, p. 183: low recomb. rate \rightarrow low polymorphism

Hitchhiking

Fig 8.20, p. 355: hitchhiking (selective sweep) example
advantageous mutation results in reduced
number of segregating sites in tightly linked region

Background selection

reduced diversity at a neutral locus
tightly surrounded by many loci of harmful mutations

Literature:

1. D.L.Hartl, A.G.Clarc. Principle of population genetics. Sinauer Associates, 2007.
2. R.Nielson, M. Statkin. An introduction to population genetics: theory and applications, Sinauer Associates. 2013.