

Estimating dN and dS

- ❖ dN - non-synonymous divergence
- ❖ dS - synonymous divergence

Nei-Gojobori (1986) Method

1. calculate number of potentially synonymous and non-synonymous sites ($s + n = 3$ per codon), disregarding stop codons

$$S = \sum_{j=1}^n \sum_{i=1}^3 f_i$$

$$N = 3n - S$$

e.g., Lysine:

AAA: $s = 1/3$

AAG: $s = 1/3$

e.g., Leucine:

TTA: $s = 2/3$

TTG: $s = 2/3$

CTA: $s = 1-1/3$

CTG: $s = 1-1/3$

CTC: $s = 1$

CTT: $s = 1$

Nei-Gojobori (1986) Method

1. calculate number of potentially synonymous and non-synonymous sites ($s + n = 3$ per codon), disregarding stop codons
2. calculate number of synonymous (s_d) and non-synonymous (n_d) differences for each codon
 1. if one difference, then obvious
 2. if two differences...

two differences

- ❖ E.g., 2 possible routes...
 - (1) TTT (Phe) - GTT (Val) - GTA (Val)
1 nonsynonymous, 1 synonymous
 - (2) TTT (Phe) - TTA (Leu) - GTA (Val)
2 nonsynonymous
- $s_d = 0.5, n_d = 1.5$

Nei-Gojobori (1986) Method

1. calculate number of potentially synonymous and non-synonymous sites ($s + n = 3$ per codon), disregarding stop codons
2. calculate number of synonymous (s_d) and non-synonymous (n_d) differences for each codon
 1. if one difference, then obvious
 2. if two differences...
 3. if three differences...

three differences

- ❖ E.g., 6 possible routes...
- (1) **TTG** (Leu) - ATG (Met) - AGG (Arg) - **AGA** (Arg)
2 nonsynonymous, 1 synonymous
 - (2) TTG (Leu) - ATG (Met) - ATA (Ile) - AGA (Arg)
 - (3) TTG (Leu) - TGG (Trp) - AGG (Arg) - AGA (Arg)
 - ~~(4) TTG (Leu) - TGG (Trp) - TGA (**Stop**) - AGA (Arg)~~
 - (5) TTG (Leu) - TTA (Leu) - ATA (Ile) - AGA (Arg)
 - ~~(6) TTG (Leu) - TTA (Leu) - TGA (**Stop**) - AGA (Arg)~~

$$s_d = 0.75, n_d = 2.25$$

Nei-Gojobori (1986) Method

$$3. dN = n_d / n,$$

$$4. dS = s_d / s$$

where n_d and s_d are the total number of synonymous and non-synonymous differences across the sequence and n and s are the average number of synonymous and nonsynonymous sites in the two sequences

	*		*	*	*	*	*	
	ACG	TAC	GTA	CGT	TTG	CCC	AAG	GAG
	Thr	Tyr	Val	Arg	Leu	Pro	Lys	Glu
s	1	1	1	1	2/3	1	1/3	1/3 = 6.33
	ACA	TAC	GTT	TGT	CTG	CCA	AGG	GAC
	Thr	Tyr	Val	Cys	Leu	Pro	Arg	Asp
s	1	1	1	1/2	4/3	1	1/3	1/3 = 6.5
s_d	1	0	1	0	1	1	0	0
n_d	0	0	0	1	0	0	1	1

$$dN = n_d / n = 3 / 17.585 = 0.171$$

$$dS = s_d / s = 4 / 6.415 = 0.624$$

$$dN/dS = 0.274$$

PAML (Phylogenetic Analysis using Maximum Likelihood)

❖ versatile program for modeling sequence evolution

- ❖ basic transition matrix

$$Q_{ij} = \begin{cases} \mu\pi_j & \text{for a synonymous transversion} \\ \mu\kappa\pi_j & \text{for a synonymous transition} \\ \mu\omega\pi_j & \text{for a nonsynonymous transversion} \\ \mu\omega\kappa\pi_j & \text{for a nonsynonymous transition} \\ 0 & \text{if } \geq 2 \text{ differences} \end{cases}$$

- ❖ estimates d_N and d_S using maximum likelihood, where κ is the transition/transversion ratio and ω is equal to d_N / d_S

$dN: dS$ ratio

- ❖ when measured in relation to the number of synonymous and non-synonymous sites
 - ❖ $dN/dS = 1$ if all substitutions are neutral
 - ❖ $dN/dS > 1$ suggests positive, diversifying selection
 - ❖ $dN/dS < 1$ suggests purifying selection (i.e., constraints on protein evolution)

Hughes & Nei 1988 Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection. *Nature* 335:167-170.

❖ examined dN / dS in human MHC class 1 genes HLA-A, B & C (n = 12 sequences)

Table 1 Mean numbers of nucleotide substitutions per 100 synonymous sites (d_S) and per 100 nonsynonymous sites (d_N)

Locus (No. sequences)	Comparisons (No.)	Antigen recognition site (ARS) (N = 57)		Remaining codons in exons 2 & 3 (N = 124, 125) [†]		Exon 4 (N = 92)	
		d_S	d_N	d_S	d_N	d_S	d_N
Human							
A (5)	vs A (10)	3.5 ± 2.0	13.3 ± 2.2***	2.5 ± 1.2	1.6 ± 0.5	9.5 ± 3.0	1.6 ± 0.7**
	vs B (20)	9.1 ± 3.3	25.1 ± 3.4***	11.9 ± 3.0	5.8 ± 0.7*	35.1 ± 8.1	2.2 ± 0.7***
	vs C (15)	7.1 ± 3.4	21.9 ± 3.5***	17.1 ± 4.0	7.5 ± 1.4*	34.9 ± 7.8	2.1 ± 1.2***
B (4)	vs B (6)	7.1 ± 3.1	18.1 ± 2.8**	6.9 ± 2.0	2.4 ± 0.7*	1.5 ± 1.1	0.5 ± 0.4
	vs C (12)	6.0 ± 2.2	22.9 ± 3.4***	14.3 ± 3.2	5.7 ± 1.1*	10.6 ± 4.0	3.1 ± 1.2
C (3)	vs C (3)	3.8 ± 2.5	8.8 ± 2.2	10.4 ± 2.8	4.8 ± 1.1	2.1 ± 1.5	1.0 ± 0.6
Overall means							
Intralocus (19)		4.7 ± 2.6	14.1 ± 2.4***	5.1 ± 2.1	2.4 ± 0.8	5.8 ± 2.0	1.1 ± 0.6**
Interlocus (47)		7.7	23.5	14.2	6.3	28.8	2.4
All comparisons (66)		6.8 ± 2.3	20.8 ± 2.3***	11.6 ± 2.1	5.2 ± 0.8**	22.1 ± 4.4	2.4 ± 0.7***
$d_S > d_N$: $d_N > d_S$		0:66		63:3		61:3†	

$dN / dS > 1$: evidence of positive, diversifying selection

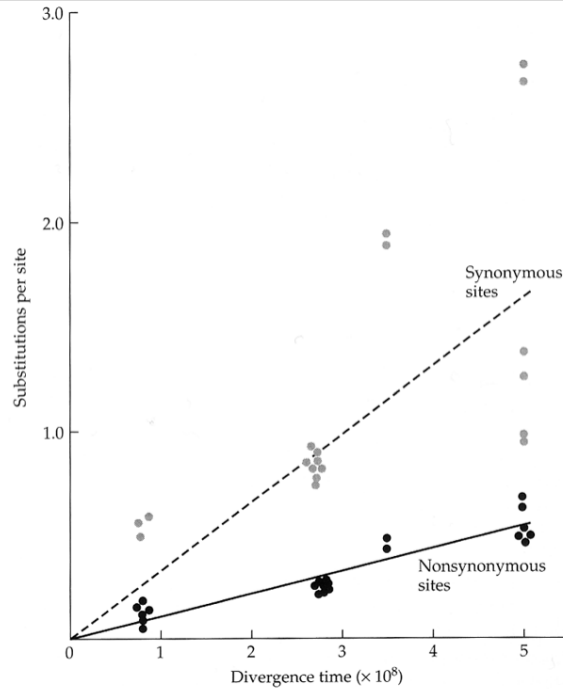
Positive Selection

- ❖ testing at the gene level is a “dull tool”
 - ❖ positive selection will usually affect one or a few codons, while the rest of the gene remains constrained ($dN/dS \ll 1$)
 - ❖ nonetheless, genes associated with immune function and reproduction (self-recognition, sperm competition, sexual conflict) often show $dN/dS > 1$
- ❖ more sophisticated methods are available to identify individual sites under selection

FIGURE 7.12 Synonymous sites and nonsynonymous sites in the β -globin gene undergo substitutions at different rates, but to a first approximation both may appear to exhibit a clocklike substitution process. (From Li et al. 1985a.)

purifying selection is the norm:

synonymous substitutions are more frequent than nonsynonymous substitutions in most genes



much greater variation among genes in non-synonymous rate than in synonymous rate

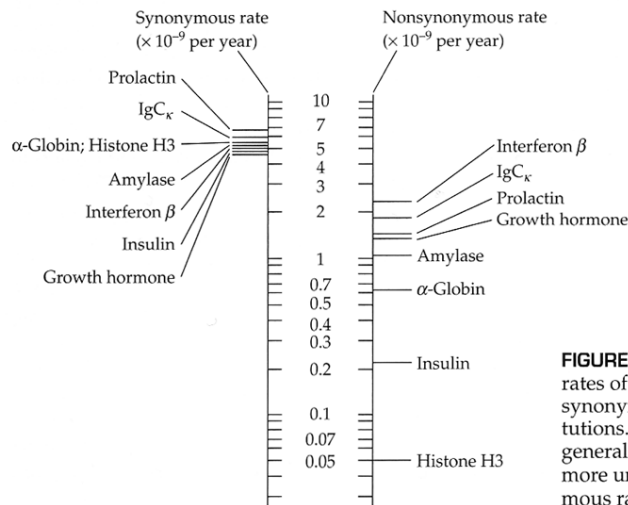
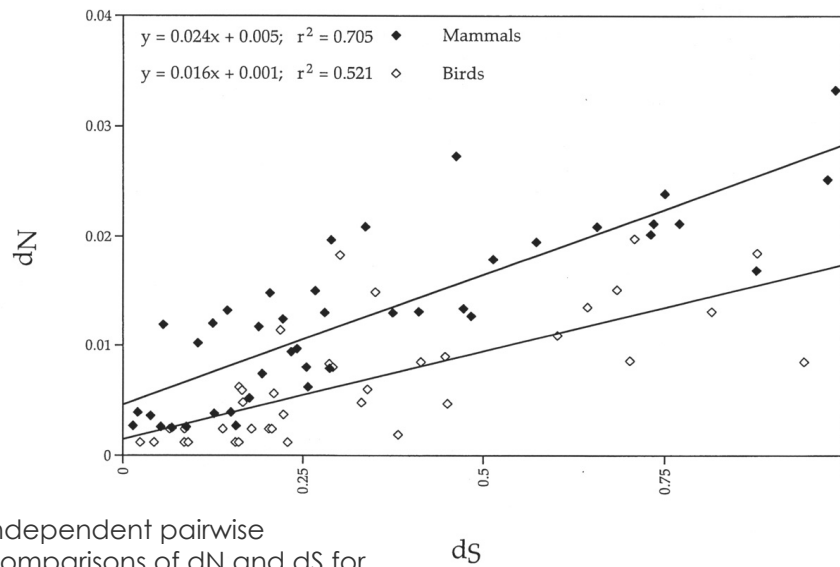


FIGURE 7.14 Comparison of rates of synonymous and nonsynonymous nucleotide substitutions. Synonymous rates are generally much faster and much more uniform than nonsynonymous rates. (From Kimura 1986.)

The avian constraint hypothesis...

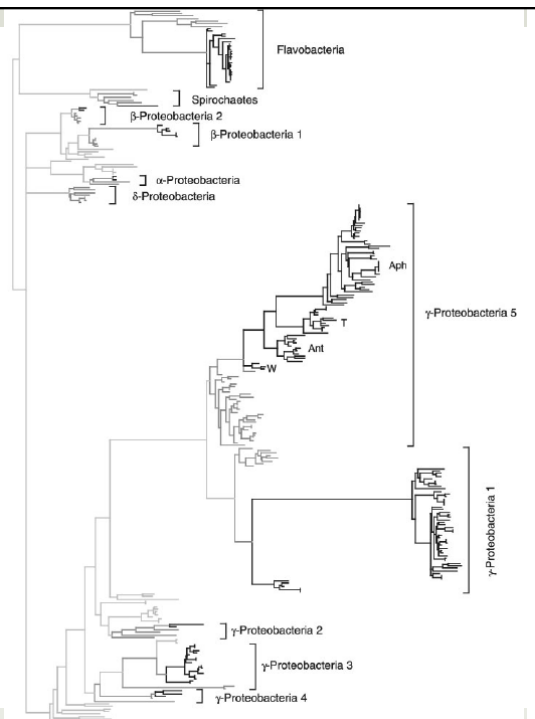


independent pairwise comparisons of dN and dS for mtDNA protein-coding genes

Stanley & Harrison MBE 1999

Woolfit & Bromham 2003
 Increased rates of sequence evolution in endosymbiotic bacteria and fungi with small effective population sizes.
MBE 20:1545-1555.

- higher rate in endosymbiotic bacteria interpreted as a consequence of nearly neutral evolution
- if so, these organisms should also have a higher dN/dS ratio...

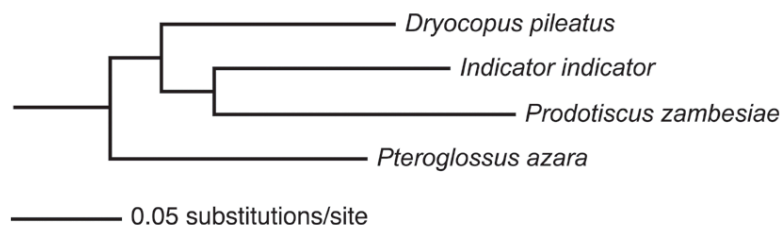


Moran (1996) Accelerated evolution and Muller's ratchet in endosymbiotic bacteria. *PNAS* 93: 2873-2878

- ❖ *Buchnera*: endosymbiont of aphids
- ❖ higher d_N/d_S than free living bacteria

Table 3. Distances based on synonymous (d_S) and nonsynonymous (d_N) substitutions for *trp* genes of two *Buchnera* and of *E. coli*-*S. typhimurium* and for *argS* of the two *Buchnera*

Gene (no. of codons)	Comparison	d_S	d_N	d_S/d_N
<i>trpa</i> (266)	<i>E. coli</i> - <i>S. typhimurium</i>	0.704 ± 0.032	0.083 ± 0.011	8.48
	<i>Buchnera</i> (Sg)- <i>Buchnera</i> (Sc)	0.570 ± 0.037	0.281 ± 0.018	2.03
<i>trpb</i> (397)	<i>E. coli</i> - <i>S. typhimurium</i>	0.581 ± 0.029	0.022 ± 0.005	26.77
	<i>Buchnera</i> (Sg)- <i>Buchnera</i> (Sc)	0.578 ± 0.031	0.167 ± 0.012	3.47
<i>trpc</i> (f)(469)	<i>E. coli</i> - <i>S. typhimurium</i>	0.629 ± 0.038	0.036 ± 0.018	17.5
	<i>Buchnera</i> (Sg)- <i>Buchnera</i> (Sc)	0.590 ± 0.035	0.273 ± 0.008	2.16
<i>trpd</i> (337)	<i>E. coli</i> - <i>S. typhimurium</i>	0.556 ± 0.031	0.016 ± 0.005	36.61
	<i>Buchnera</i> (Sg)- <i>Buchnera</i> (Sc)	0.567 ± 0.034	0.269 ± 0.016	2.11
<i>trpe</i> (520)	<i>E. coli</i> - <i>S. typhimurium</i>	0.577 ± 0.025	0.071 ± 0.008	8.13
	<i>Buchnera</i> (Sg)- <i>Buchnera</i> (Sc)	0.689 ± 0.026	0.264 ± 0.013	2.61
<i>argS</i> (131)	<i>Buchnera</i> (Sg)- <i>Buchnera</i> (Sc)	0.533 ± 0.056	0.370 ± 0.029	1.44



$$dN/dS = 0.0409$$

Indicator



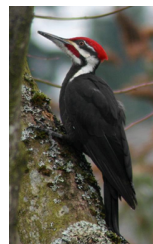
Prodotiscus



Brood Parasites

$$dN/dS = 0.0256$$

Dryocopus



Pteroglossus



