

Nei-Gojobori dN/dS estimate (NG86)

Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions

M. Nei and T. Gojobori

Mol. Biol. Evol. **3** 418–426 (1986)

- For each codon C we define $ES(C)$ and $EN(C)$ - the numbers of synonymous and non-synonymous *sites* of a codon
 - e.g., $ES(GAA) = 1/3$, $EN(GAA) = 8/3$.
- May also define them as fractions of substitutions that do not lead to stop codons,
 - e.g., $ES(GAA) = 1/3$, $EN(GAA) = 7/3$.
- The sum of ES and EN over all codons in a sequence gives an estimate of expected synonymous and non-synonymous **sites** in a sequence.
- For two sequences (the target of the original method), we average $ES(C)$ and $EN(C)$ at each site.
- EN/ES is thus the ***expected ratio of non-synonymous to synonymous substitutions counts under neutral evolution***

NG86 example

Mean	Seq1	ACA	ATA	ATC	TTT	AAT	CAA
	Syn	1	2/3	2/3	1/3	1/3	1/3
	NonSyn	2	7/3	7/3	8/3	8/3	7/3
	Seq2	ACA	ATA	ACC	TTT	AAC	CAA
	Syn	1	2/3	1	1/3	1/3	1/3
	NonSyn	2	7/3	2	8/3	8/3	7/3
	Syn	1	2/3	5/6	1/3	1/3	1/3
	NonSyn	2	7/3	13/6	8/3	8/3	7/3

ES = $3\frac{1}{2}$, EN = $14\frac{1}{6}$: under neutrality, we expect the ratio of non-synonymous to synonymous substitutions of **EN/ES** ~ 4.05