NG86 limitations: averaging across all sites in a gene

- Different sites in a gene will be subject to different selective forces.

- A *gene-wide* measure of selection is going to average these effects.

- **Most** sites in **most** genes will be maintained by purifying selection.

- Positively selected sites are of great biological interest, because they point to how a particular gene can respond to selective pressures.

- Negatively selected sites are also of interest, because they point to functional constraint, and could be used to guide drug or vaccine design.

- Must develop methods that are able to disentangle the contributions of individual sites.

Suzuki-Gojobori (SG99): the penultimate extension of NG86

Uses a tree to compute dN/dS at a given site

1. Reconstruct ancestral sequences by nucleotide-level <u>parsimony</u>

2. Compute `EN` and `ES` using labeled branches; define p_e = `ES/EN`

3. Compute `S` and `NS` for each site (minimum evolution)

4. Estimate the probability that the number of synonymous substitutions `S` is unusually low (positive selection) or unusually high (negative selection), using the binomial distribution given p_e from step 2.

A method for detecting positive selection at single amino acid sites
Y. Suzuki and T. Gojobori
Mol Biol Evol 16 1315-1328 (1999)

>500 citations