Codon-substitution models

- In 1994, first tractable mechanistic evolutionary models for codon sequences were proposed by **Muse and Gaut** (MG94), and, independently, by **Goldman and Yang** (GY94) [in the same issue of MBE, back to back]

- Markov models of codon substitution provide a powerful framework for **estimating substitution rates** from coding sequence data, as they

  - *encode our mechanistic understanding of the evolutionary process,*

  - *enable one to compute the phylogenetic likelihood,*

  - *permit hypothesis testing or Bayesian inference,*

  - *systematically account for confounding processes (unequal base frequencies, nucleotide substitution biases, etc.),*

  - *afford many opportunities for extension and refinement (still happening today).*

A likelihood approach for comparing synonymous and nonsynonymous nucleotide substitution rates, with application to the chloroplast genome
S. V. Muse and B. S. Gaut
Mol Biol Evol 11 715-724 (1994)

~1000 citations

A codon-based model of nucleotide substitution for protein-coding DNA sequences.
N. Goldman and Z. Yang
Mol Biol Evol 11 725--736 (1994)

~2250 citations

Rate matrix for an MG-style codon model

$$
(\text{Rate})_{X,Y}(dt) =
\begin{cases}
\textcolor{green}{\alpha} & \pi_t dt & , & \text{one-step, synonymous substitution,} \\
\textcolor{red}{\beta} & \pi_t dt & , & \text{one-step, non-synonymous substitution,} \\
0 & & , & \text{multi-step.}
\end{cases}
$$

X,Y = AAA...TTT (excluding stop codons),
π_t - frequency of the target nucleotide.

Example substitutions:
AAC→AAT (one step, synonymous - Asparagine)
CAC→GAC (one step, non-synonymous - Histidine to Aspartic Acid)
AAC→GTC (multi-step).

αR_{CT}
βR_{CG}

α (syn. rate) and **β** (non-syn. rate) are the key quantities for all selection analyses