# 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)**

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

# Three example datasets

- **West Nile Virus NS3 protein**
  - An interesting case study of how positive selection detection methods lead to testable hypotheses for function discovery
  - Brault et al 2007, A single positively selected West Nile viral mutation confers increased virogenesis in American crows
- **HIV-1 transmission pair**
  - Partial *env* sequences from two epidemiologically linked individuals
  - An example of multiple selective environments (source, recipient, transmission)
- **SARS-CoV-2 Spike**
  - Full length spike sequences chosen to represent viral diversity (circa mid 2021)
  - Good example for analyzing selection in population samples with many “dead-end” intra-host variants