Limitations: Multiple substitutions

- **The model assumes that point mutations alter one nucleotide at a time**, hence most of the instantaneous rates:
  - (`3134/3761` or `84.2%` in the case of the universal genetic code) are `0`. (Sparse)
- This restriction, however, does not mean that the model disallows any substitutions that involve multiple nucleotides (e.g., **`ACT`**⟹**`AGG`**).
  - This can be further relaxed with models supporting multiple nucleotide changes.
- Such substitutions must simply be realized via several single nucleotide steps, e.g., **`ACT`**⟹**`AGT`**⟹**`AGG`**
- In fact the `(i,j)` element of **`T(t) = exp(Qt)`** sums the probabilities of all such possible pathways of duration **t**, including reversions
- Compare this to the naive NG86 parsimony approach.

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