Algebraic Methods in Phylogenetics

Marta Casanellas Rius

Dept. Matemàtiques i Institut de Matemàtiques de la UPC-BarcelonaTech (IMTech)

Universitat Politècnica de Catalunya

Centre de Recerca Matemàtica

BGSMath Summer school

Juliol 2022

General Markov model (GMM)

- π distribution at root: no constraints, 3 free parameters
- P: 12 free parameters per edge
- The most general model (assuming i.i.d)

$$M_i = \begin{array}{c} \\ A \\ C \\ G \\ T \end{array} \begin{array}{c} \begin{array}{cccc} A & C & G & T \end{array} \\ \begin{pmatrix} a_i & b_i & c_i & d_i \\ e_i & f_i & g_i & h_i \\ j_i & k_i & l_i & m_i \\ n_i & o_i & p_i & q_i \end{pmatrix} \end{array}$$

Jukes-Cantor model (JC69)

- π uniform distribution: $\pi_{\mathtt{A}} = ... = \pi_{\mathtt{T}} = 0.25$
- M: 1 free parameter per edge

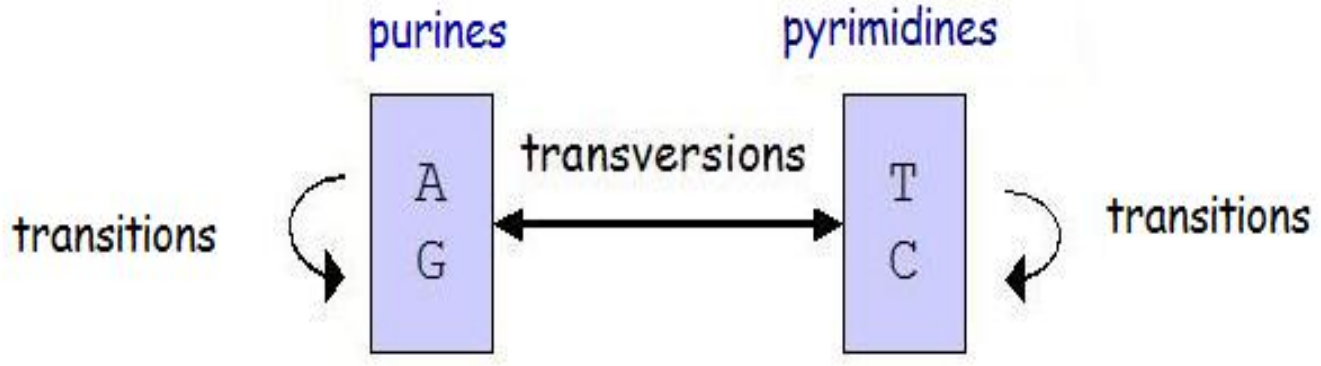
$$M_i = \begin{array}{c} \\ A \\ C \\ G \\ T \end{array} \begin{array}{c} \begin{array}{cccc} A & C & G & T \end{array} \\ \begin{pmatrix} a_i & b_i & b_i & b_i \\ b_i & a_i & b_i & b_i \\ b_i & b_i & a_i & b_i \\ b_i & b_i & b_i & a_i \end{pmatrix} \end{array}, \quad a_i + 3b_i = 1$$

Kimura 3-parameter model (K81)

- π uniform distribution: $\pi_{\mathtt{A}} = ... = \pi_{\mathtt{T}} = 0.25$
- M: 3 free parameters per edge

$$M_i = \begin{array}{c} \\ A \\ C \\ G \\ T \end{array} \begin{array}{c} \begin{array}{cccc} A & C & G & T \end{array} \\ \begin{pmatrix} a_i & b_i & c_i & d_i \\ b_i & a_i & d_i & c_i \\ c_i & d_i & a_i & b_i \\ d_i & c_i & b_i & a_i \end{pmatrix} \end{array}, \quad a_i + b_i + c_i + d_i = 1$$

- **Kimura 2-parameter (K80)**: $b_i = d_i$ (2 free parameters per edge)



Strand symmetric model (SSM)

- π satisfies: $\pi_{\mathtt{A}} = \pi_{\mathtt{T}}, \pi_{\mathtt{C}} = \pi_{\mathtt{G}}$
- M: 6 free parameters per edge,

$$P_i = \begin{array}{c} \\ A \\ C \\ G \\ T \end{array} \begin{array}{c} \begin{array}{cccc} A & C & G & T \end{array} \\ \begin{pmatrix} a_i & b_i & c_i & d_i \\ e_i & f_i & g_i & h_i \\ h_i & g_i & f_i & e_i \\ d_i & c_i & b_i & a_i \end{pmatrix} \end{array}$$

Strand symmetry in a DNA molecule:

A–T

C–G