

# PEMARF – aula 8

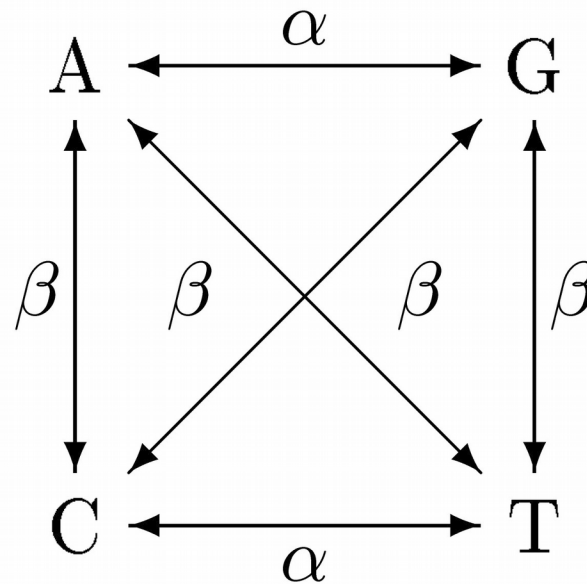
- Modelos
- Discussão
- Intervalo
- Seminário
- Correção
- Ensaios

# Modelos e Realidade

- “All models are wrong, but some are useful”

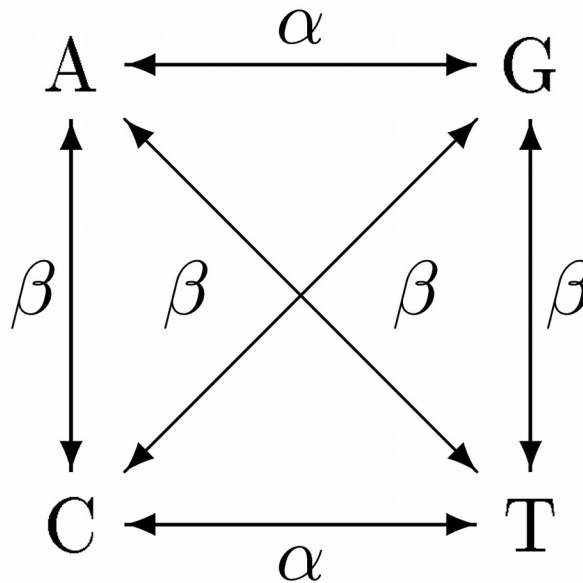
# Modelos

- Estimar N parâmetros de substituição



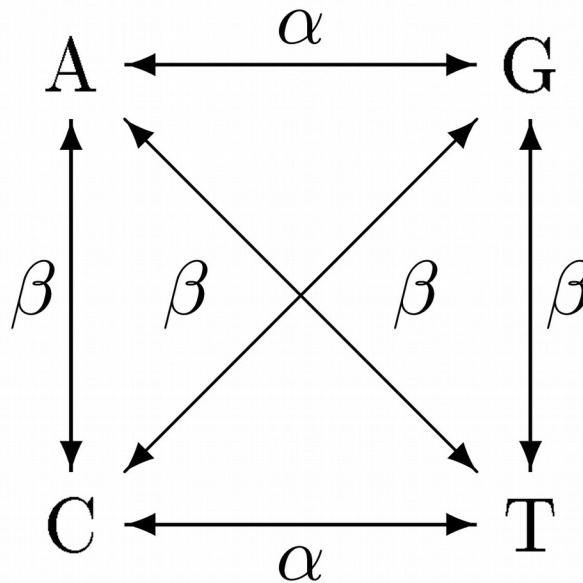
# Modelos

- Estimar N parâmetros de substituição
- Simples – JC
- Complexo – GTR
- “aninhamento”



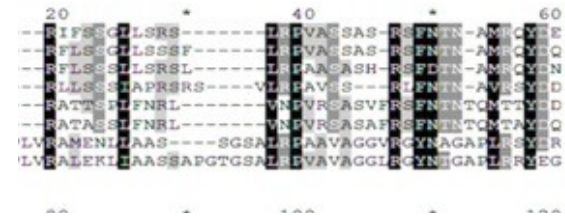
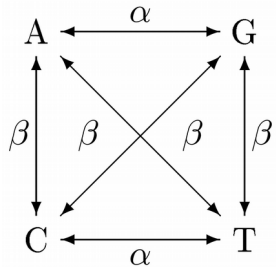
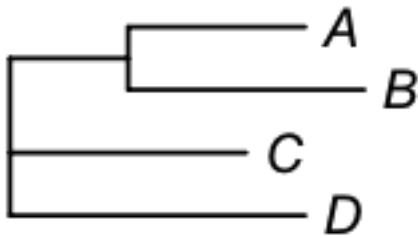
# Modelos

- Frequência de bases (F)
- Sítios inváriaveis (I)
- Variação de taxas entre-sítios (Gamma)



# Modelos

- Estimativa – sobre árvores
- Árvores-base



# Escolha do Modelo

- LRT
  - $\Delta = 2(\ln L_1 - \ln L_0)$
  - distribuição de  $\Delta$ ?
- Critério de Informação (AIC)
  - $AIC_i = -2 \ln L_i + 2N_i$

# Escolha do modelo

- LRT – hierárquica ou dinâmica

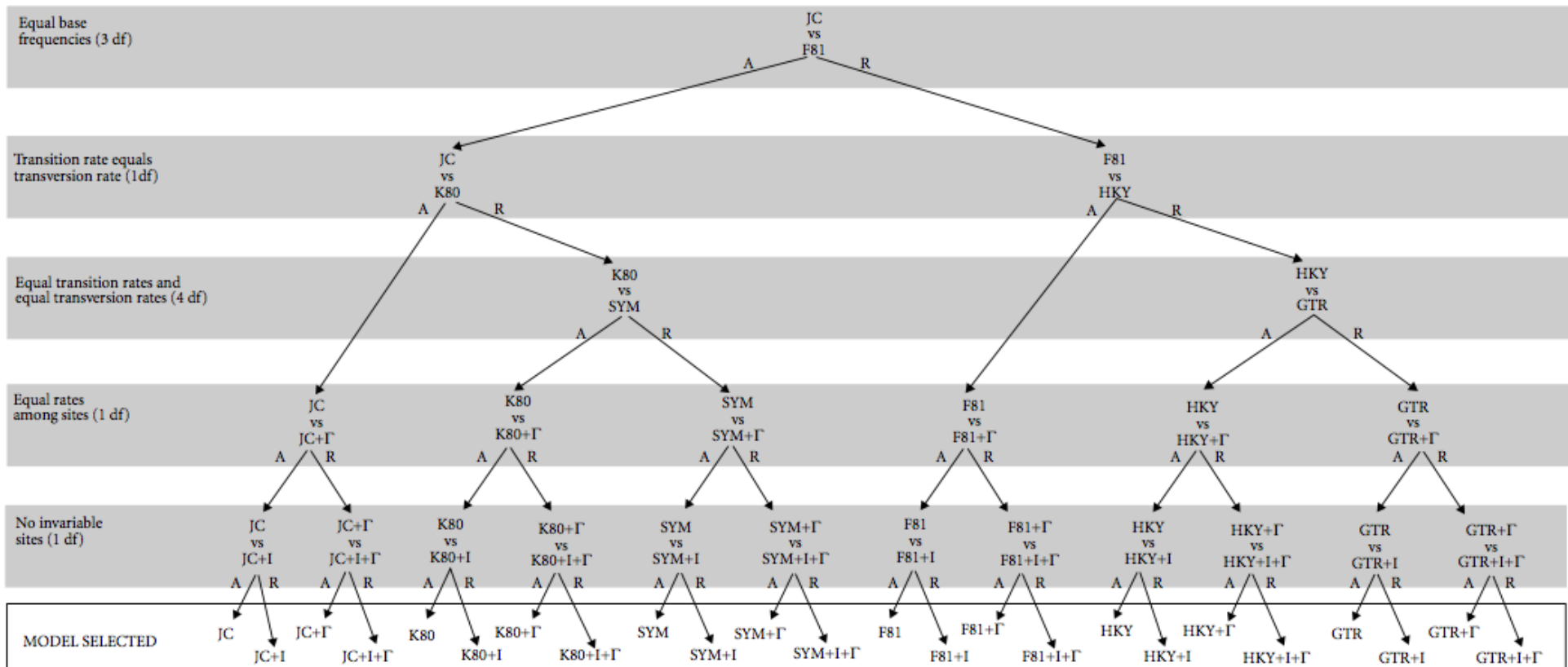


Figure 10.1 A comparison of models of nucleotide substitution. Model-selection methods selected the best-fit model for the data set at hand among 24 possible models. The models of DNA substitution are JC (Jukes and Cantor, 1969), K80 (Kimura, 1980), SYM (Zharkikh, 1994), F81 (Felsenstein, 1981), HKY (Hasegawa et al., 1985), and GTR (Rodríguez et al., 1990).  $\Gamma$ : rate heterogeneity among sites; I: proportion of invariable sites; df: degrees of freedom.



# Escolha do modelo

- LRT – hierárquica ou dinâmica

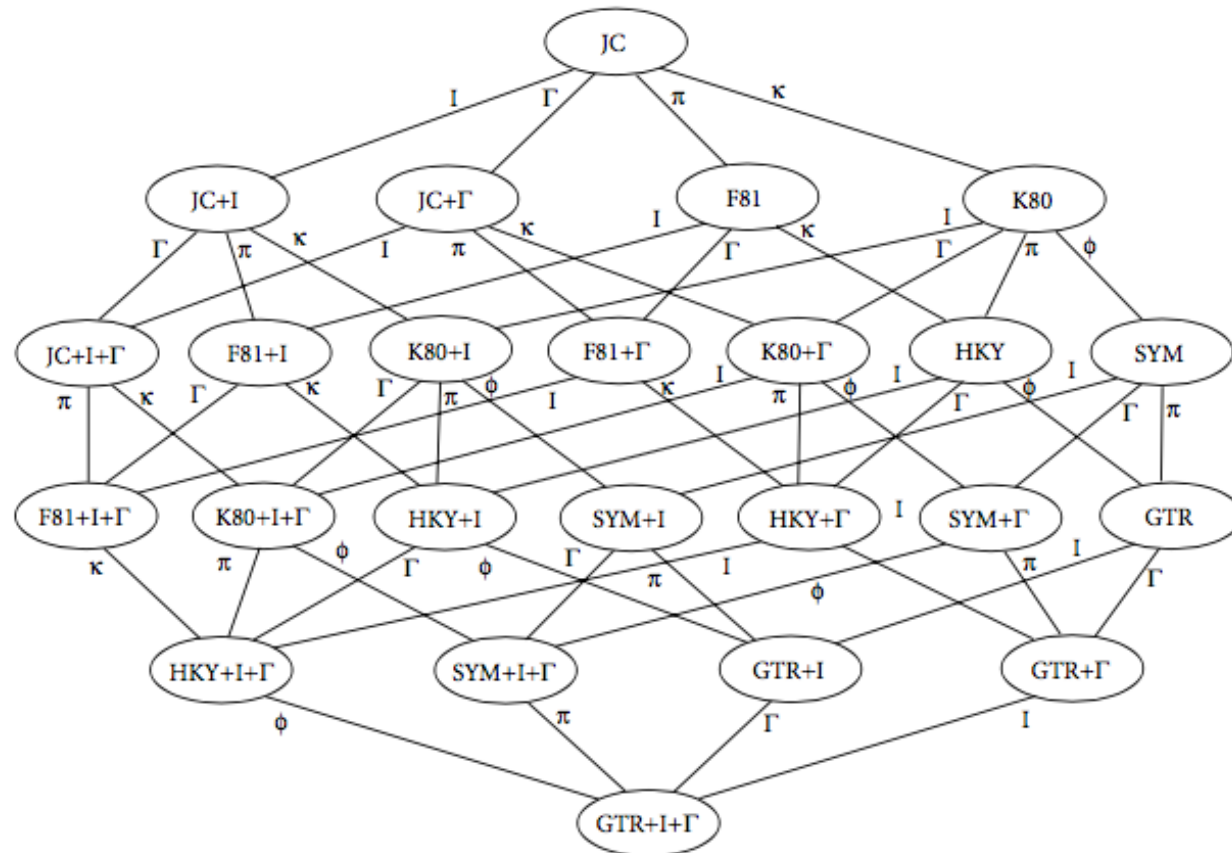


Figure 10.2 Dynamic LRTs. Starting with the simplest (JC) or the most complex model (GTR+I+ $\Gamma$ ), LRTs are performed among the current model and the alternative models that maximize the difference in likelihood.  $\pi$ : base frequencies;  $\kappa$ : transition/transversion bias;  $\phi$ : substitution rates among nucleotides;  $\Gamma$ : rate heterogeneity among sites;  $I$ : proportion of invariable sites.

# Escolha do Modelo

- AIC, BIC – comparar todos
- AICc, BICc – corrigidos para populações pequenas (modelos de nucleotídeos)

# Modelos de Amino-ácidos

- Potencialmente muito mais complexos:
  - nt – cerca de 24 modelos (já incluindo F, G, I)
  - AA – combinatorial de 20x20 \* cada um combinado com F, G, I
  - tipicamente, são consideradas cerca de 130 possibilidades
  - abordagem empírica
  - Le & Gascuel (2009) - LG