Estimation example

Input:
• Alignment.
• Model parameters from neutral sequence

\[ P(x_t = T \mid x_0 = G) = [e^{Qt}]_{TG} \]
Estimation example 2
Different gene conservation patterns

Protein Coding Gene

Known lincRNA- XIST

RepA
Estimating $\pi$

Decompose $Q$ by “extracting” the stationary distribution:

$$Q_{ab} = R_{ab} \pi_b$$

$R$: Neutral substitution pattern

$\pi$: Site specific forces

Find a ML estimator for $\pi$ using the EM algorithm.

Score:

$$\Delta L = \log(\mathcal{L}(\pi)) - \log(\mathcal{L}(\pi_0))$$
This column is not conforming to the rules!
Proof of concept

$\omega$

$\pi$

43% vs 16% detection by $\pi$ vs. $\omega$
A generalization: Conserved motif discovery

Find regions in the alignment whose substitution pattern is explained by the motif.

$$S_{\text{Cons}}(\text{alignment}) = \log \left( \frac{P(\text{align\_col}_1|\pi_1^M) \cdots P(\text{align\_col}_{10}|\pi_{10}^M)}{P(\text{align\_col}_1|\pi_N^N) \cdots \log(P(\text{align\_col}_{10}|\pi_N^N))} \right)$$
P53 Motif instance conservation

P53

MDM2

Novel non coding gene

M. Huarte, O. Zuk, M. Guttmann