Supplementary Figure 1

a

H $H_{\bullet,j}$

$g =$ acagtcaccgtcga $= H_{0,\bullet}$

$a^1 =$ agagt--ccgccgg $= H_{1,\bullet}$

$a^2 =$ acagt--ccgtcgg $= H_{2,\bullet}$

* ***   *** **

L

b

t_2 g

v_2

t_1 t_3

v_1 a^1

t_4

a^2

c

$H_{\bullet,j}$

t_2 t

v_2

t_1 t_3

v_1 c

t_4

t

1) $P_X(v_\delta \rightarrow v_{\delta'} | t_e) = exp(Xt_e)$

2) $P_X(H_{\bullet,j}) = \sum_V (P_X(v_1)P_X(v_1 \rightarrow v_2|t_1)P_X(v_2 \rightarrow t|t_2)^* \\ P_X(v_2 \rightarrow c|t_3)P_X(v_1 \rightarrow t|t_4))$

3) $\varphi = log_2 \left(\dfrac{P_{MEHB}(H_{\bullet,j})}{P_{neutral}(H_{\bullet,j})} \right)$

d

$w_j \rightarrow \Xi(w_j)$

m

φ/m

overlapping motif matches

$\xi(j)$

j

L

e

t_2 ..acaata..

v_2

t_1 t_3

v_1 alignment gap ..-------..

t_4

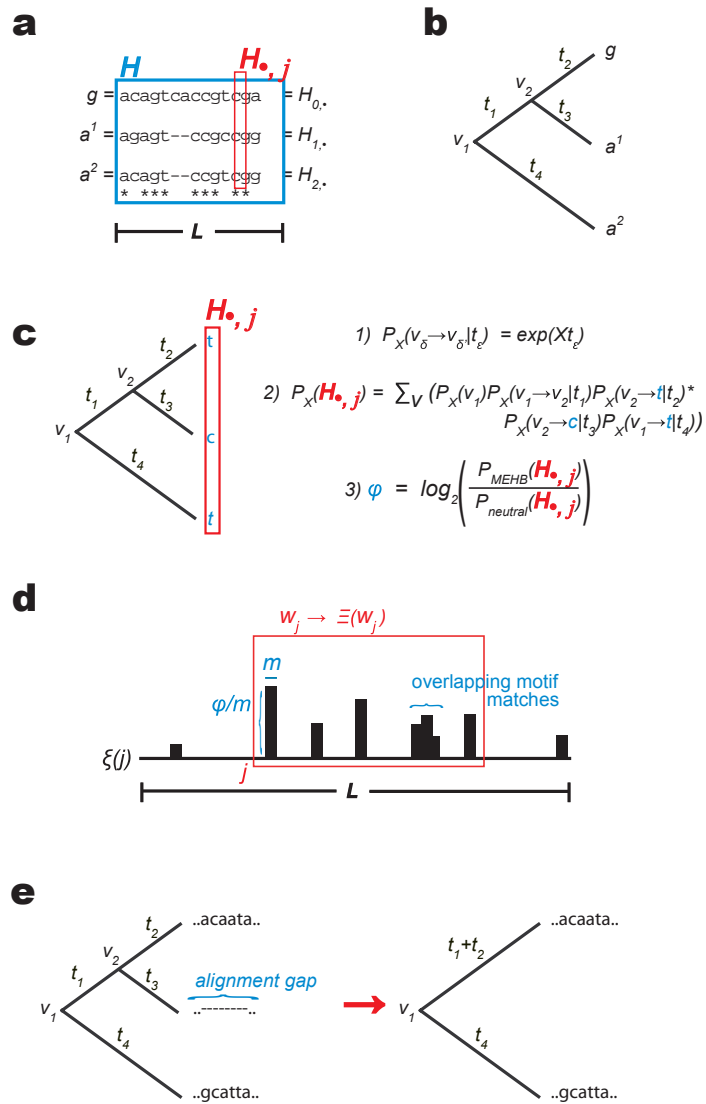..gcatta..

$\rightarrow$

..acaata..

t_1+t_2

v_1

t_4

..gcatta..

Supplementary Figure 1: PhylCRM scoring scheme for a single motif
(**a**) g represents the sequence being searched for CRMs and a^1 and a^2 are sequences from another organism aligned to it. L represents the length of the sequence, $H_{0,\bullet} = g$, $H_{i,\bullet} = a^{(i)}$, and $H_{\bullet,j}$ denotes the alignment column at position j.
(**b**) Tree indicating the phylogeny of g, a^1, and a^2. (**c**) Scoring motif matches using the MEHB model. Here, the probability that a given nucleotide a turns into b during time t is given by a matrix exponential, for a suitably chosen rate-matrix R. This probability is then used to compute the probability of observing the set of nucleotides $H_{\bullet,j}$ under both the MEHB rate-matrix and the neutral matrix. The score of the motif φ is then taken to be the log-likelihood of the ratio of these probabilities. (**d**) Graphical image of scores for a motif M along g, where the height of the bars is φ/m. These scores are stored in an array ξ and the score of a window w_j (represented by $\Xi(w_j)$) is then given by summing ξ in w_j. (**e**) When there is no alignable sequence at a given position (or if there is no motif match there), the branch containing that sequence is removed and the pruned tree is used to compute φ.