## PSSM's with pseudocounts

Local multiple alignment
ungapped

*k* sequences
*w* sites

Pseudocounts
*b* = 0.1

| | w=10 | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| k=11 | G | T | C | A | G | G | T | G | T | G |
| | A | T | C | A | A | G | T | G | C | A |
| | A | C | C | A | C | C | T | G | C | A |
| | A | T | C | A | A | G | T | G | C | T |
| | G | G | C | A | G | G | T | G | T | C |
| | A | G | C | A | G | G | T | G | T | G |
| | G | G | C | A | G | T | T | G | A | A |
| | G | G | C | A | A | C | T | G | G | A |
| | G | T | C | A | G | G | T | G | G | C |
| | G | A | C | A | G | A | T | G | G | G |
| | G | T | C | A | A | G | T | G | G | G |
| \|Σ\|=4 | A | A | A | A | A | A | A | A | A | A |
| | G | G | G | G | G | G | G | G | G | G |
| | C | C | C | C | C | C | C | C | C | C |
| | T | T | T | T | T | T | T | T | T | T |

Nucleotide counts
*c[x, i] + b* :
Number of copies of nucleotide *x* in column *i* .

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| A | 4.1 | 1.1 | 0.1 | 11.1 | 4.1 | 1.1 | 0.1 | 0.1 | 1.1 | 4.1 |
| G | 7.1 | 4.1 | 0.1 | 0.1 | 6.1 | 7.1 | 0.1 | 11.1 | 4.1 | 4.1 |
| C | 0.1 | 1.1 | 11.1 | 0.1 | 1.1 | 2.1 | 0.1 | 0.1 | 3.1 | 2.1 |
| T | 0.1 | 5.1 | 0.1 | 0.1 | 0.1 | 1.1 | 11.1 | 0.1 | 3.1 | 1.1 |

Frequency matrix

$q[x, i] = \frac{c[x, i] + b}{k + |\Sigma| b}$

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| A | 0.36 | 0.10 | 0.01 | 0.97 | 0.36 | 0.10 | 0.01 | 0.01 | 0.10 | 0.36 |
| G | 0.62 | 0.36 | 0.01 | 0.01 | 0.54 | 0.62 | 0.01 | 0.97 | 0.36 | 0.36 |
| C | 0.01 | 0.10 | 0.97 | 0.01 | 0.10 | 0.18 | 0.01 | 0.01 | 0.27 | 0.18 |
| T | 0.01 | 0.45 | 0.01 | 0.01 | 0.01 | 0.10 | 0.97 | 0.01 | 0.27 | 0.10 |

$k + |\Sigma| b = 11 + 4 \cdot 0.1$
Frequency of nucleotide x in column *i* , corrected with pseudocount *b*.

Likelihood ratio

$P[x, i] = \frac{q[x, i]}{p[x]}$

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| A | 1.22 | 0.33 | 0.03 | 3.30 | 1.22 | 0.33 | 0.03 | 0.03 | 0.33 | 1.22 |
| G | 3.04 | 1.75 | 0.04 | 0.04 | 2.61 | 3.04 | 0.04 | 4.75 | 1.75 | 1.75 |
| C | 0.04 | 0.47 | 4.75 | 0.04 | 0.47 | 0.90 | 0.04 | 0.04 | 1.33 | 0.90 |
| T | 0.03 | 1.52 | 0.03 | 0.03 | 0.03 | 0.33 | 3.30 | 0.03 | 0.92 | 0.33 |

Background frequency

| | |
|---|---|
| A | 0.295 |
| G | 0.205 |
| C | 0.205 |
| T | 0.295 |

PSSM: Log-odds scoring matrix
*S[x, i] = log_2 P[x,i]*

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| A | 0.29 | -1.61 | -5.07 | 1.72 | 0.29 | -1.61 | -5.07 | -5.07 | -1.61 | 0.29 |
| G | 1.60 | 0.81 | -4.55 | -4.55 | 1.38 | 1.60 | -4.55 | 2.25 | 0.81 | 0.81 |
| C | -4.55 | -1.09 | 2.25 | -4.55 | -1.09 | -0.15 | -4.55 | -4.55 | 0.41 | -0.15 |
| T | -5.07 | 0.60 | -5.07 | -5.07 | -5.07 | -1.61 | 1.72 | -5.07 | -0.12 | -1.61 |

Log-odds score of observing nucleotide x in column *i*.

## Scoring a new sequence

PSSM: Log-odds scoring matrix

S[x, i] = log_2 P[x,i]

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| A | 0.29 | -1.61 | -5.07 | 1.72 | 0.29 | -1.61 | -5.07 | -5.07 | -1.61 | 0.29 |
| G | 1.60 | 0.81 | -4.55 | -4.55 | 1.38 | 1.60 | -4.55 | 2.25 | 0.81 | 0.81 |
| C | -4.55 | -1.09 | 2.25 | -4.55 | -1.09 | -0.15 | -4.55 | -4.55 | 0.41 | -0.15 |
| T | -5.07 | 0.60 | -5.07 | -5.07 | -5.07 | -1.61 | 1.72 | -5.07 | -0.12 | -1.61 |

| | T | C | C | A | A | C | T | G | A | C | A | G | A | T | G | G | G |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0.15 | -5.07 | -1.09 | 2.25 | 1.72 | 0.29 | -0.15 | 1.72 | 2.25 | -1.61 | -0.15 | | | | | | | |
| -20.61 | | -4.55 | -1.09 | -5.07 | 1.72 | -1.09 | -1.61 | -4.55 | -5.07 | 0.41 | 0.29 | | | | | | |
| -29.67 | | | -4.55 | -1.61 | -5.07 | -4.55 | -5.07 | 1.60 | -5.07 | -4.55 | -1.61 | 0.81 | | | | | |
| -12.90 | | | | 0.29 | -1.61 | 2.25 | -5.07 | 1.38 | -1.61 | -4.55 | -5.07 | 0.81 | 0.29 | | | | |
| -16.34 | | | | | 0.29 | -1.09 | -5.07 | -4.55 | 0.29 | -0.15 | -5.07 | 2.25 | -1.61 | -1.61 | | | |
| -18.39 | | | | | | -4.55 | 0.60 | -4.55 | 1.72 | -1.09 | -1.61 | -4.55 | -5.07 | -0.12 | 0.81 | | |
| -20.51 | | | | | | | -5.07 | 0.81 | -5.07 | -4.55 | 0.29 | 1.60 | -5.07 | -5.07 | 0.81 | 0.81 | |
| 9.33 | | | | | | | | 1.60 | -1.61 | 2.25 | 1.72 | 1.38 | -1.61 | 1.72 | 2.25 | 0.81 | 0.81 |