

Nucleotide Frequencies Matrices for TSS motif from various sets of plant promoters*

Nucleotide Frequencies Matrix for TSS motif from 236 experimentally verified dicot plant TATA promoters

	-4	-3	-2	-1	+1	+2	+3	+4
A	0.318	0.186	0.127	0.085	0.928	0.258	0.318	0.445
C	0.212	0.309	0.161	0.737	0.017	0.242	0.390	0.237
G	0.076	0.089	0.081	0.059	0.038	0.127	0.102	0.102
T	0.394	0.415	0.631	0.119	0.017	0.373	0.191	0.216
	w	y	t	C	A	h	m	h

Nucleotide Frequencies Matrix for TSS motif from 121 experimentally verified dicot plant TATA-less promoters

	-4	-3	-2	-1	+1	+2	+3	+4
A	0.455	0.264	0.107	0.157	0.579	0.446	0.149	0.322
C	0.091	0.256	0.033	0.595	0.231	0.058	0.314	0.124
G	0.025	0.099	0.033	0.174	0.041	0.074	0.140	0.099
T	0.430	0.380	0.826	0.074	0.149	0.421	0.397	0.455
	w	h	T	c	m	w	y	w

Nucleotide Frequencies Matrix for TSS motif from 356 experimentally verified dicot plant promoters

	-4	-3	-2	-1	+1	+2	+3	+4
A	0.343	0.214	0.131	0.109	0.866	0.291	0.277	0.417
C	0.180	0.271	0.131	0.714	0.049	0.180	0.360	0.183
G	0.066	0.094	0.063	0.089	0.029	0.129	0.114	0.100
T	0.411	0.420	0.674	0.089	0.057	0.400	0.249	0.300
	w	h	T	C	A	w	h	w

Nucleotide Frequencies Matrix for TSS motif from 107 experimentally verified monocot plant promoters

	-4	-3	-2	-1	+1	+2	+3	+4
A	0.131	0.168	0.477	0.262	0.140	0.009	0.972	0.112
C	0.467	0.346	0.280	0.308	0.430	0.916	0.009	0.383
G	0.140	0.093	0.084	0.159	0.187	0.019	0.000	0.140
T	0.262	0.393	0.159	0.271	0.243	0.056	0.019	0.364

y	y	m	n	y	C	A	y
---	---	---	---	---	---	---	---

* +1 corresponds to the annotated TSS.

The following IUPAC codes for nucleotide abbreviations are used:

A – Adenine, C – Cytosine, G – Guanine, T – Thymine, Y – C or T, S – G or C, W – A or T,

M – A or C, H – A or C or T, N – any base (<http://www.bioinformatics.org/sms/iupac.html>).

To get consensus sequences, the following thresholds were applied:

- Single nucleotide in lower case – if its frequency >0.5 , but <0.65 ;
- Single nucleotide in upper case – if its frequency ≥ 0.65 ;
- Combination of two nucleotides in lower case – if a sum of their frequencies ≥ 0.65 , but <0.9 ;
- Combination of two nucleotides in upper case – if a sum of their frequencies >0.9 ;
- Combination of three nucleotides in lower case – if a sum of their frequencies >0.85 ;
- n – other cases.