

摘要：搜索基因组中 CpxR 潜在结合位点并用 pssm 打分。

生成 pwm (Position-Weight Matrices) 数据源：15bp 数据见表格 1，16bp 将第 8 个碱基重复一次作为数据源。

生成 pssm (Position-Specific Scoring Matrices)：pseudocounts=0.5, log2, 假定四种碱基背景概率相等。

PSSM 信息：

15bp:

mean = 11.43, standard deviation = 4.22															
max = 19.53, min = -42.31															
threshold = 7.22															
	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
A:	-4.43	-4.43	1.88	1.84	1.46	0.70	0.43	0.62	0.33	-0.73	-1.62	-0.52	1.64	1.80	1.30
C:	-4.43	-0.97	-2.10	-4.43	-0.73	-1.62	0.43	-0.52	-0.52	0.53	-2.84	-0.73	-0.97	-2.84	-1.26
G:	1.95	-2.84	-4.43	-2.84	-1.62	0.22	-0.97	-0.73	0.53	-0.18	1.72	-4.43	-4.43	-4.43	-0.18
T:	-4.43	1.72	-4.43	-2.10	-1.62	-0.18	-0.34	0.22	-0.73	0.10	-2.10	1.41	-1.62	-1.62	-2.10

16bp:

mean = 11.53, standard deviation = 4.25																
max = 20.15, min = -43.04																
threshold = 7.29																
	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
A:	-4.43	-4.43	1.88	1.84	1.46	0.70	0.43	0.62	0.62	0.33	-0.73	-1.62	-0.52	1.64	1.80	1.30
C:	-4.43	-0.97	-2.10	-4.43	-0.73	-1.62	0.43	-0.52	-0.52	-0.52	0.53	-2.84	-0.73	-0.97	-2.84	-1.26
G:	1.95	-2.84	-4.43	-2.84	-1.62	0.22	-0.97	-0.73	-0.73	0.53	-0.18	1.72	-4.43	-4.43	-4.43	-0.18
T:	-4.43	1.72	-4.43	-2.10	-1.62	-0.18	-0.34	0.22	0.22	-0.73	0.10	-2.10	1.41	-1.62	-1.62	-2.10

mean 即相对熵

Score threshold: 越小序列越多，实际意义：2^x 倍的背景概率

Searching for matches using the PSSM score:

基因组: NC_000913.3 (W3110) CP009273.1(BW25113)

结果说明: pos 代表 cpxR 结合第一个碱基位置，“_c”代表互补链，seq 为序列。Score 即以上打分，默认按照序列位置排序，这样容易发现一个基因存在多个潜在结合位点的情况。也可以按照 score 自行排序，也可给定其它筛选阈值我来重新给出结果。Location 代表潜在结合位点所在位置的基因信息。

tail 列说明:

NULL: 基因间

gene short than 200bp

left 100bp

right 100bp

注意: 所谓左右没有管基因方向，只是从基因组序列的 5'到 3'

表格 1: CpxR 结合位点数据源

注: ppiD 以及 ldtC (ycfS)序列更长，无法纳入打分。

acrD	GTATAAAAGGGCATG
aer	GTAACAGAGTGAAAG
alx (ygjT)	GTAAAACAGTGTAAG
aroG	GTAAAACCCCGTTTA
bacA	GTAAAGCGGCGAAAA
bamE	GCAAAACGCCGTAAG
cpxR/cpxP1	GTAAAACAACGTAAA
cpxR/cpxP2	GCAAACATGCGTCAG
csgB1	GTAACGCAGCGTTAA
csgB2	GTAAAAAATTGTCCA
csgD1	GTAATAATGTAAAC
degP1 (htrA)	gtaaagacgaacaat
degP2	gtaaattaccgtcag
dgcZ (ydeH)	gtcaatatttcata
dsbC	GTAAAGGCAACAAAG
efeU*(ycdN)	ggaatgtttataaaa
fabA	gtagaagaaggcaaa
fabB	gtaaggctgcgcaaa
ftnB	Gtaaaaatatataaa
ldtD	GCAACTCCCTGAAAC
marR	gcaagtataagtcaa
motA	GTAAAAAGACGTAAA
mviM	GTAAATCACCGTCAA
mzrA (yqjA)	Gtcagcctgtgtaaa
ompC	GTAAAGAAGGGTAAA
ompF1	GTAAACAAAAGTAAA
ompF2	GTAATATTCCGTAAAC
ppiA	Gtaaaattaggtaaa
psd	GTAAAAACGCGTAAA
rseD	GTAAAGTTCCGTAAAC
skp	GTAAAACCTGGTAAG
slt	GTAACGTGGCGTAAA
spy	gtaaaacaatgaaag
srkA (yihE)	gtaaaagcttgtaag
tomB (ybaJ)	GTAATTTTTCGTAAT
tsr	GTAAAGTAAGGTAAA
ung	GCAAAAAAATGTAAA
yccA	gtaaagatgggtaaa
yebE	gtatatcgagacaaa
yidQ	GTAAAGCGGGGATAA
yqaE	GTAAATGAGAGTAAA