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（另外增加两个16bp真实数据）。

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 = 12.04, standard deviation = 4.22

max = 19.78, min = -44.53

threshold = 7.82

0 1 2 3 4 5 6 7 8 9 10 11 12 13 14

A: -4.64 -4.64 1.90 1.86 1.46 0.57 0.31 0.64 0.40 -0.74 -1.84 -0.74 1.70 1.83 1.42

C: -4.64 -0.94 -2.32 -4.64 -0.74 -1.84 0.40 -0.40 -0.56 0.64 -3.06 -0.56 -1.18 -3.06 -1.47

G: 1.96 -3.06 -4.64 -3.06 -1.47 0.31 -0.56 -0.94 0.49 -0.40 1.77 -4.64 -4.64 -4.64 -0.40

T: -4.64 1.73 -4.64 -2.32 -1.84 0.00 -0.40 0.21 -0.74 0.11 -2.32 1.42 -1.84 -1.84 -2.32

16bp：

mean = 12.15, standard deviation = 4.25

max = 20.43, min = -45.48

threshold = 7.90

0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15

A: -4.64 -4.64 1.90 1.86 1.46 0.57 0.31 0.64 0.64 0.40 -0.74 -1.84 -0.74 1.70 1.83 1.42

C: -4.64 -0.94 -2.32 -4.64 -0.74 -1.84 0.40 -0.40 -0.40 -0.56 0.64 -3.06 -0.56 -1.18 -3.06 -1.47

G: 1.96 -3.06 -4.64 -3.06 -1.47 0.31 -0.56 -0.94 -0.94 0.49 -0.40 1.77 -4.64 -4.64 -4.64 -0.40

T: -4.64 1.73 -4.64 -2.32 -1.84 0.00 -0.40 0.21 0.21 -0.74 0.11 -2.32 1.42 -1.84 -1.84 -2.32

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： 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(15bp)**

注：ppiD以及ldtC (ycfS)序列更长，无法纳入打分。

|  |  |
| --- | --- |
| acrD | GTATAAAAGGGCATG |
| aer | GTAACAGAGTGAAAG |
| alx (ygjT) | GTAAAACAGTGTAAA |
| aroG | GTAAAACCCCGTTTA |
| bacA | GTAAAGCGGCGAAAA |
| bamE | GCAAAACGCCGTAAG |
| cpxR/cpxP1 | GTAAAACAACGTAAA |
| cpxR/cpxP2 | GCAAACATGCGTCAG |
| csgB1 | GTAACGCAGCGTTAA |
| csgB2 | GTAAAAAATTGTCCA |
| csgD1 | GTAACTAAATGTAAC |
| degP1 (htrA) | gtaaagacgaacaat |
| degP2 | gtaaattaccgtcag |
| dgcZ (ydeH) | gtcaatattttcata |
| 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 | GTAACCAAAAGTAAA |
| ompF2 | GTAATATTCCGTAAC |
| ppiA | Gtaaaattaggtaaa |
| psd | GTAAAAACGCGTAAA |
| rseD | GTAAAGTTCCGTAAC |
| skp | GTAAAACCTGGTAAG |
| slt | GTAACGTGGCGTAAA |
| spy | gtaaaacaatgaaag |
| srkA (yihE) | gtaaaagcttgtaag |
| tomB (ybaJ) | GTAATTTTTCGTAAT |
| tsr | GTAAAGTAAGGTAAA |
| ung | GCAAAAAAATGTAAA |
| yccA | gtaaagatgggtaaa |
| yebE | gtatatcgagacaaa |
| yidQ | GTAAAGCGGGGATAA |
| yqaE | GTAAATGAGAGTAAA |
| carA | GTAAATCAATGCAAA |
| dusB | GCAAATTTGCGTAAA |
| shoB | GTAAAGCACAGCAAA |
| chaA | GTAAAAATTCGTAAA |
| ampH | GTAACGGCGCGTAAA |
| amiC | GTAAAGGAACGTAAA |
| acs | GTAAGTGCATGTAAA |

红色：赵喆新增，amiC为文献报道。

**16bp数据源：**

所有15bp数据源将中间碱基重复，新增：

|  |  |
| --- | --- |
| xylE | GTAAACGCATTGTAAA |
| ldtC (ycfS) | gcaaagaatttgtaaa |

XylE：赵喆新增，ldtC上次16bp未使用。