# psi-blast工具（pssm特征获取）

PSI-BLAST (Position-Specific Iterative Basic Local Alignment Search Tool) 是一种更加高灵敏的BLASTp程序，对于发现远亲物种的相似蛋白或某个蛋白家族的新成员非常有效。通过 PSI-BLAST工具我们可以获得pssm打分矩阵。

Position Specific Scoring Matrix(PSSM)是蛋白质BLAST搜索中使用的一种评分矩阵，序列某些区域中较高的PSSM得分通常表明该序列与PSSM表征的家族或基序可能存在生物学关系 。它可用于预测残基溶剂的可及性[4]，蛋白质二级结构[5]，残基-残基接触图[6]，蛋白质无序区[7]，蛋白质结合位点[8]，蛋白质-DNA相互作用[9]或蛋白质-蛋白质界面热点[10]。PSSM分数通常显示为正整数或负整数。正值表示给定的氨基酸取代比偶然发生的频率更高，而负值表示该取代发生的频率低于预期。大的阳性分数通常表明关键的功能残基，可能是活性位点残基或其他分子间相互作用所需的残基。

我们从三方面介绍：1、**软件安装和数据库下载** 2、**软件使用参数说明** 3、**计算结果说明**

## 1、软件安装

软件下载链接为：https://blast.ncbi.nlm.nih.gov/Blast.cgi?CMD=Web&PAGE\_TYPE=BlastNews#1，直接下载链接为：ftp://ftp.ncbi.nlm.nih.gov/blast/executables/blast+/，选择对应的版本进行下载。此处我的系统为linux系统，下载的版本为：ncbi-blast-2.6.0+-x64-linux.tar.gz

### 1.1、解压文件夹：

tar -zxvf ncbi-blast-2.6.0+-x64-linux.tar.gz

### 1.2、配置环境变量

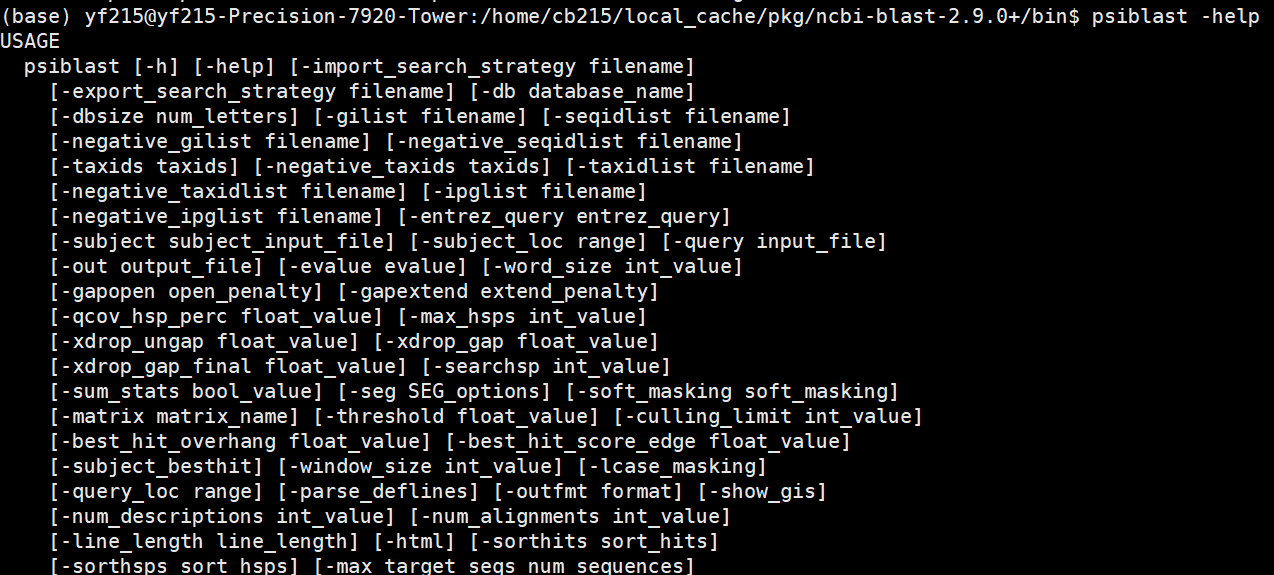
此处我的安装路径为：/home/cb215/local\_cache/pkg/ncbi-blast-2.6.0+/bin 向文件 ~/.bashrc 中写入:

export PATH=/home/cb215/local\_cache/pkg/ncbi-blast-2.6.0+/bin:${PATH}

保存~/.bashrc并退出，然后通过 source ~/.bashrc 激活环境变量

### 1.3、验证程序是否安装成功

输入：psiblast -help 得到如下结果说明安装成功



windows安装教程：https://blog.csdn.net/u013313168/article/details/77508367

## 2、软件使用参数说明

psiblast -help查看所有参数的详细信息，在此，仅对几个常用的参数进行说明，其他参数请自行查询。

**-query** 要查询的蛋白质文件，必须是fasta格式，例如：

>1AKHA  
KKEKSPKGKSSISPQARAFLEEVFRRKQSLNSKEKEEVAKKCGITPLQVRVWFINKRMRSK

**-num\_iterations** 迭代次数 （整数类型一般会放3）

**-db** 后面放你格式化的数据库名称 如：nr

**-out** 你希望输出文件名，可任意取名

**-out\_ascii\_pssm** 如果你需要生成pssm矩阵，则需要输入此参数，此参数后放pssm矩阵的名字

**-evalue** 阈值E，默认是10

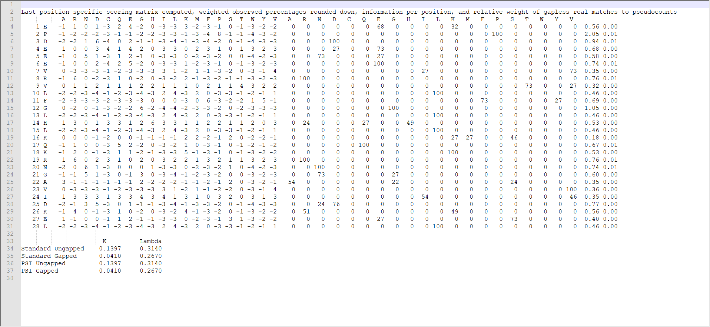
例如：

psiblast -db /home/cb215/local\_cache/pkg/database/blast+/nr\_db/nr\_db -query test.fasta -num\_iterations 3 -evalue 0.001 -out\_ascii\_pssm test.pssm

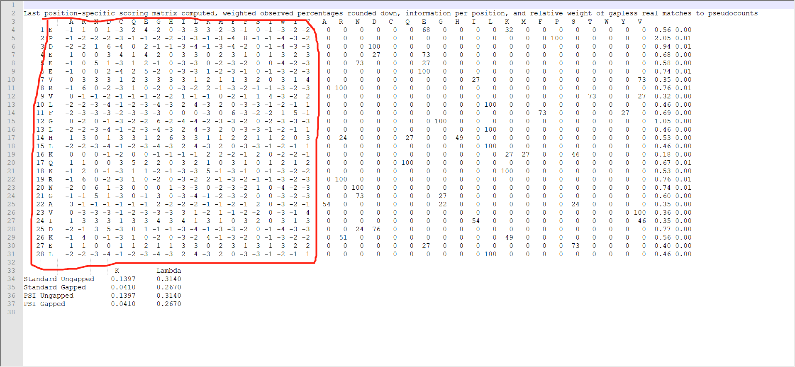
其中数据库文件使用的是nr数据库，数据下载链接为：ftp://ftp.ncbi.nlm.nih.gov/blast/db/，可自行选择数据库

## 3、计算结果说明

结果前半部分是打分矩阵，后半部分是每个位置对应的残基比例 。运行上述代码得到结果如下：



这里我们需要的PSSM矩阵就为L\*20（L为所使用的fasta序列的长度，20个氨基酸），如下图中红圈所示：



## 参考材料

1. 3DCONS-DB: A Database of Position-Specific Scoring Matrices in Protein Structures

2. https://www.cnblogs.com/cong3Z/p/12775414.html

3. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 1997 Sep 1;25(17):3389-402. doi: 10.1093/nar/25.17.3389. PMID: 9254694; PMCID: PMC146917.

4. Yang, Y.; Heffernan, R.; Paliwal, K.; Lyons, J.; Dehzangi, A.; Sharma, A.; Wang, J.; Sattar, A.; Zhou, Y. Spider2: A package to predict secondary structure, accessible surface area, and main-chain torsional angles by deep neural networks. Methods Mol. Biol. 2017, 1484, 55–63. [Google Scholar] [PubMed]

5. Wang, S.; Peng, J.; Ma, J.; Xu, J. Protein secondary structure prediction using deep convolutional neural fields. Sci. Rep. 2016, 6, 18962. [Google Scholar] [CrossRef] [PubMed]

6. Skwark, M.J.; Raimondi, D.; Michel, M.; Elofsson, A. Improved contact predictions using the recognition of protein like contact patterns. PLoS Comput. Biol. 2014, 10, e1003889. [Google Scholar] [CrossRef] [PubMed]

7. Ishida, T.; Kinoshita, K. Prdos: Prediction of disordered protein regions from amino acid sequence. Nucleic Acids Res. 2007, 35, W460–W464. [Google Scholar] [CrossRef] [PubMed]

8. Zhou, J.; Xu, R.; He, Y.; Lu, Q.; Wang, H.; Kong, B. Pdnasite: Identification of DNA-binding site from protein sequence by incorporating spatial and sequence context. Sci. Rep. 2016, 6, 27653. [Google Scholar] [CrossRef] [PubMed]

9. Wei, L.; Tang, J.; Zou, Q. Local-DPP: An improved DNA-binding protein prediction method by exploring local evolutionary information. Inf. Sci. 2017, 384, 135–144. [Google Scholar] [CrossRef]

10. Melo, R.; Fieldhouse, R.; Melo, A.; Correia, J.D.; Cordeiro, M.N.; Gumus, Z.H.; Costa, J.; Bonvin, A.M.; Moreira, I.S. A machine learning approach for hot-spot detection at protein-protein interfaces. Int. J. Mol. Sci. 2016, 17. [Google Scholar] [CrossRef] [PubMed]