
“Linux 生物信息基础”课程
小组集体练习、讨论、交流

总 结 报 告

组：4 次：4 组长：陈奕晗 执笔：陈奕晗

1. 时间：2021 年 4 月 15 日，15:00 ~ 17:00
2. 地点：王克桢 348
3. 人员：陈奕晗、邹济平、朱瑾煜、高培翔
4. 方式：线下讨论
5. 主题：
 - 1) BLAST 数据库相似性搜索
 - 2) 本小组综合课题选题方向

6. 内容

6.1 BLAST 数据库相似性搜索

6.1.1 NCBI BLAST

(1) 搜索方式：

blastp 蛋白—蛋白

blastn 核酸—核酸

blastx 核酸—蛋白

tblastn 蛋白—核酸

tblastx 编码核酸—编码核酸

(2) 例题

1. 以人血红蛋白 alpha 亚基 (HBA_HUMAN) 为检测序列, 用 BlastP 搜索 Swiss-Prot 数据库中 12 个珠蛋白

Standard Protein BLAST

Enter Query Sequence

Enter accession number(s), gis, or FASTA sequence(s) [Clear](#) [Query subrange](#)

From To

Or, upload file [选择文件](#) [未选择文件](#)

Job Title

☐ Align two or more sequences

Choose Search Set

Database [UniProtKB/Swiss-Prot \(swissprot\)](#)

Organism [human \(taxid:9606\)](#) [exclude](#) [Add organism](#)

Exclude ☐ Models (XMP) ☐ Non-redundant RefSeq proteins (WP) ☐ Uncultured/environmental sample sequences

Program Selection

Algorithm ☒ blastp (protein-protein BLAST)

☐ PSI-BLAST (Position-Specific Iterated BLAST)

☐ PHI-BLAST (Pattern Hit Initiated BLAST)

☐ DELTA-BLAST (Domain Enhanced Lookup Time Accelerated BLAST)

Choose a BLAST algorithm

[BLAST](#) Search database swissprot using Blastp (protein-protein BLAST)

☐ Show results in a new window

Algorithm parameters

General Parameters

Max target sequences

Short queries ☒ Automatically adjust parameters for short input sequences

Expect threshold

Word size

Max matches in a query range

Scoring Parameters

Matrix

Gap Costs

Compositional adjustments

Filters and Masking

Filter ☐ Low complexity regions

Mask ☐ Mask for lookup table only

☐ Mask lower case letters

[BLAST](#) Search database swissprot using Blastp (protein-protein BLAST)

☐ Show results in a new window

参数设置:

目标序列:	P69905
选择搜索的数据库:	UniProtKB/Swiss-Prot (swissprot)
选择搜索物种:	human (taxid:9606)
program selection (Algorithm):	blastp (protein-protein BLAST)
Max target sequences:	100
word size:	2 (wordsize 越小, 灵敏度越高)
Matrix:	PAM250 (PAM250 的灵敏度比 BLOSUM62 等其他选项都高)
Compositional adjustments:	Composition-based statistics

搜索结果

Sequences producing significant alignments

[Download](#) [Select columns](#) [Show](#) [100](#)

☒ select all 12 sequences selected [GenPept](#) [Graphics](#) [Distance tree of results](#) [Multiple alignment](#) [MSA Viewer](#)

Description	Scientific Name	Max Score	Total Score	Query E-value	Percent Identity	Accession
BackName: Full-Hemoglobin subunit alpha; AltName: Full-Alpha-globin; AltName: Full-Hemoglobin alpha chain; Homo sapiens	Homo sapiens	174	174	100%	142	P69905.2
BackName: Full-Hemoglobin subunit beta-1; AltName: Full-Hemoglobin beta-1 chain; AltName: Full-Beta-1-globin; Homo sapiens	Homo sapiens	131	131	100%	61.97%	P02106.2
BackName: Full-Hemoglobin subunit beta; AltName: Full-Beta-2-globin; AltName: Full-Hemoglobin beta chain; AltName: Full-Beta-2-globin; Homo sapiens	Homo sapiens	129	129	100%	59.96%	P02108.2
BackName: Full-Hemoglobin subunit gamma-1; AltName: Full-Hemoglobin gamma-1 chain; AltName: Full-Gamma-1-globin; Homo sapiens	Homo sapiens	109	109	99%	56.33%	Q69909.1
BackName: Full-Hemoglobin subunit delta; AltName: Full-Delta-globin; AltName: Full-Hemoglobin delta chain; Homo sapiens	Homo sapiens	88.0	88.0	98%	46.24%	P08871.2
BackName: Full-Hemoglobin subunit gamma-2; AltName: Full-Gamma-2-globin; AltName: Full-Beta-2-globin; Homo sapiens	Homo sapiens	86.3	86.3	98%	26.23%	P08882.2
BackName: Full-Hemoglobin subunit delta; AltName: Full-Delta-globin; AltName: Full-Hemoglobin delta chain; Homo sapiens	Homo sapiens	88.0	88.0	98%	36.23%	P02104.2
BackName: Full-Hemoglobin subunit gamma-1; AltName: Full-Gamma-1-globin; AltName: Full-Beta-2-globin; Homo sapiens	Homo sapiens	85.7	85.7	98%	36.23%	P08881.2
BackName: Full-Hemoglobin subunit epsilon; AltName: Full-Epsilon-globin; AltName: Full-Hemoglobin epsilon chain; Homo sapiens	Homo sapiens	81.0	81.0	98%	36.23%	P02102.2
BackName: Full-Hemoglobin subunit gamma-2; AltName: Full-Gamma-2-globin; AltName: Full-Beta-2-globin; Homo sapiens	Homo sapiens	82.9	82.9	98%	16.13%	Q69909.1
BackName: Full-Hemoglobin subunit gamma-1; AltName: Full-Gamma-1-globin; AltName: Full-Beta-2-globin; Homo sapiens	Homo sapiens	55.5	55.5	100%	56.11%	P02144.2
BackName: Full-Hemoglobin subunit gamma-2; AltName: Full-Gamma-2-globin; AltName: Full-Beta-2-globin; Homo sapiens	Homo sapiens	32.2	32.2	95%	0.03%	Q69909.1

点击 Multiple alignment 可调用多序列比对程序

Alignments [Select All](#) [Re-align](#) [Mouse over the sequence identifier for sequence title](#)

View Format: [Compact](#) [Conservation](#) [Setting](#) [2 Bits](#)

✓ P69905.2	1	MY-LSPADKTVKAKGKYGARAGYAEALRMPFSPTTKTFP-HF-----DLSPSSAQVGRGKDYADLT	68
✓ P02106.2	1	MS-LTKETITLYSMKISTQDITSTELRFLSPQTKTFP-HF-----DLHPSAQGRGKDYAVG	68
✓ P02104.2	1	MYLTPKTSATVNLGKYYN---DAVGEALGRLLVYPTGRFF-SFGLSSPDVWMPKYGKDYLGAFS	73
✓ P08881.2	1	MGHTEEDKATITSLGKYYN---EDAGETLGRLLVYPTGRFF-SFGLSSPDVWMPKYGKDYLGAFS	73
✓ P09109.2	1	MA-LSAEDKALYRLKLGKYYVITTEALRFLPAKTKTFP-HL-----DLSPSSAQVGRGKDYADLT	68
✓ P08871.2	1	MGHTEEDKATITSLGKYYN---EDAGETLGRLLVYPTGRFF-SFGLSSPDVWMPKYGKDYLGAFS	73
✓ P02108.2	1	MYLTPKTSATVNLGKYYN---DAVGEALGRLLVYPTGRFF-SFGLSSPDVWMPKYGKDYLGAFS	73
✓ P02102.2	1	MYHTAEKAAVTSYGKYYN---EDAGETLGRLLVYPTGRFF-SFGLSSPDVWMPKYGKDYLGAFS	73
✓ Q69909.1	1	---HLSAQRAQIAQVVDLIGRRAQFASLLRLFTVPTSTKYFP-HLSA-----CQATGLLRSGRKLAVG	67
✓ P02144.2	1	---HLSAQRAQIAQVVDLIGRRAQFASLLRLFTVPTSTKYFP-HLSA-----CQATGLLRSGRKLAVG	67
✓ Q69909.1	1	[15]SEELSEAGRAVQVHARLTKCEPHVATLVYFPVPSAQVPS-QPKMDEPDESGRLSGRKLAVG	90
✓ Q69909.1	1	---HNPFPFLINGSRVSRKPLRNTLFLARLALRPLLPQV-NKQPSFSDCLSSPFLNTRKYLVD	73
✓ P69905.2	69	NATVHY---DQMPKYL---LSALSGRAHRLRYDPYDFELLSKCLLYLAHLPAFTFVYASLQPLASVSTLTSTKTR	142
✓ P02106.2	69	DATKESL---DQI644---LSALSGRAHRLRYDPYDFELLSKCLLYLAHLPAFTFVYASLQPLASVSTLTSTKTR	142
✓ P02104.2	74	DGLAHL---DQI647---FQGLSELKDKLYDPENFELLSKCLLYLAHLPAFTFVYASLQPLASVSTLTSTKTR	147
✓ P08881.2	74	DGLAHL---DQI647---FQGLSELKDKLYDPENFELLSKCLLYLAHLPAFTFVYASLQPLASVSTLTSTKTR	147
✓ P09109.2	74	LATKHL---DQI648---LSALSGRAHRLRYDPYDFELLSKCLLYLAHLPAFTFVYASLQPLASVSTLTSTKTR	142
✓ P08881.2	74	DGLAHL---DQI647---FQGLSELKDKLYDPENFELLSKCLLYLAHLPAFTFVYASLQPLASVSTLTSTKTR	147
✓ P08871.2	74	DGLAHL---DQI647---FQGLSELKDKLYDPENFELLSKCLLYLAHLPAFTFVYASLQPLASVSTLTSTKTR	147
✓ P02102.2	74	DGLAHL---DQI647---FQGLSELKDKLYDPENFELLSKCLLYLAHLPAFTFVYASLQPLASVSTLTSTKTR	147
✓ Q69909.1	74	DGLAHL---DQI647---FQGLSELKDKLYDPENFELLSKCLLYLAHLPAFTFVYASLQPLASVSTLTSTKTR	147
✓ P02144.2	75	GILKKKth---HEAL---TEPLAGRAHRLRYDPYDFELLSKCLLYLAHLPAFTFVYASLQPLASVSTLTSTKTR	154
✓ Q69909.1	91	TVYENL-HQPKYSGLVLSALSGRAHRLRYDPYDFELLSKCLLYLAHLPAFTFVYASLQPLASVSTLTSTKTR	190
✓ Q69909.1	74	ANTVHY---DQI648---LSALSGRAHRLRYDPYDFELLSKCLLYLAHLPAFTFVYASLQPLASVSTLTSTKTR	151

2. 以人血红蛋白 alpha 亚基（HBA_HUMAN）为检测序列，用 PSI-Blast 搜索 Swiss-Prot 数据库，找出人珠蛋白家族成员脑红蛋白（Neuroglobin）。

第一次搜索

Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession	Select for PSI blast	Used to build PSSM	New add
RecName: Full-Hemoglobin subunit alpha; AltName: Full-Alpha-globin; AltName: Full-Hemo...	Homo sapiens	256	256	100%	5e-81	100.00%	142	P00905.2	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	
RecName: Full-Hemoglobin subunit theta-1; AltName: Full-Hemoglobin theta-1 chain; AltName...	Homo sapiens	183	183	100%	5e-43	81.67%	142	P00905.2	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	
RecName: Full-Hemoglobin subunit zeta; AltName: Full-HBAZ; AltName: Full-Hemoglobin ze...	Homo sapiens	178	178	100%	4e-81	59.86%	142	P02008.2	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	
RecName: Full-Hemoglobin subunit mu; AltName: Full-Hemoglobin mu chain; AltName: Full...	Homo sapiens	150	150	99%	2e-49	45.39%	141	Q68K93.1	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	
RecName: Full-Hemoglobin subunit gamma-1; AltName: Full-Gamma-1-globin; AltName: Full...	Homo sapiens	117	117	98%	5e-36	40.41%	147	P08891.2	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	
RecName: Full-Hemoglobin subunit gamma-2; AltName: Full-Gamma-2-globin; AltName: Full...	Homo sapiens	117	117	98%	5e-36	40.41%	147	P08892.2	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	
RecName: Full-Hemoglobin subunit beta; AltName: Full-Beta-globin; AltName: Full-Hemo...	Homo sapiens	117	117	98%	7e-36	43.15%	147	P08871.2	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	
RecName: Full-Hemoglobin subunit delta; AltName: Full-Delta-globin; AltName: Full-Hemo...	Homo sapiens	114	114	98%	9e-35	43.15%	147	P02042.2	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	
RecName: Full-Hemoglobin subunit epsilon; AltName: Full-Epsilon-globin; AltName: Full-H...	Homo sapiens	110	110	98%	2e-33	36.99%	147	P02100.2	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	
RecName: Full-Cytoglobin; AltName: Full-Histoglobin; Short-Hb; AltName: Full-Slate; cell...	Homo sapiens	171.1	171.1	96%	3e-19	28.08%	190	Q9WMA5.1	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	
RecName: Full-Myoglobin (Homo sapiens)	Homo sapiens	62.0	62.0	100%	1e-13	27.52%	154	P02144.2	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	

二次迭代搜索找到脑红蛋白 Neuroglobin

Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession	Select for PSI blast	Used to build PSSM	New add
RecName: Full-Hemoglobin subunit zeta; AltName: Full-HBAZ; AltName: Full-Hemoglobin ze...	Homo sapiens	207	207	100%	4e-73	59.86%	142	P02008.2	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	
RecName: Full-Hemoglobin subunit alpha; AltName: Full-Alpha-globin; AltName: Full-Hemo...	Homo sapiens	196	196	100%	2e-69	100.00%	142	P00905.2	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	
RecName: Full-Hemoglobin subunit delta; AltName: Full-Delta-globin; AltName: Full-Hemo...	Homo sapiens	196	196	98%	2e-68	41.78%	147	P02042.2	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	
RecName: Full-Hemoglobin subunit gamma-2; AltName: Full-Gamma-2-globin; AltName: Full...	Homo sapiens	196	196	98%	3e-68	40.41%	147	P08892.2	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	
RecName: Full-Hemoglobin subunit theta-1; AltName: Full-Hemoglobin theta-1 chain; AltName...	Homo sapiens	194	194	100%	1e-67	81.67%	142	P00905.2	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	
RecName: Full-Hemoglobin subunit gamma-1; AltName: Full-Gamma-1-globin; AltName: Full...	Homo sapiens	194	194	98%	2e-67	40.41%	147	P08891.2	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	
RecName: Full-Hemoglobin subunit beta; AltName: Full-Beta-globin; AltName: Full-Hemo...	Homo sapiens	193	193	98%	3e-67	41.78%	147	P08871.2	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	
RecName: Full-Hemoglobin subunit epsilon; AltName: Full-Epsilon-globin; AltName: Full-H...	Homo sapiens	190	190	98%	9e-66	36.99%	147	P02100.2	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	
RecName: Full-Hemoglobin subunit mu; AltName: Full-Hemoglobin mu chain; AltName: Full...	Homo sapiens	188	188	99%	2e-65	45.39%	141	Q68K93.1	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	
RecName: Full-Myoglobin (Homo sapiens)	Homo sapiens	169	169	100%	5e-57	26.35%	154	P02144.2	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	
RecName: Full-Cytoglobin; AltName: Full-Histoglobin; Short-Hb; AltName: Full-Slate; cell...	Homo sapiens	170	170	96%	7e-57	28.08%	190	Q9WMA5.1	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	
RecName: Full-Neuroglobin (Homo sapiens)	Homo sapiens	58.1	58.1	92%	2e-11	22.54%	151	Q9NPG2.1	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	

PSI-blast 为一种序列相似性搜索算法，可多次迭代得到更多结果

3. 以人血红蛋白 alpha 亚基(HBA_HUMAN)为检测序列，搜索 Swiss-Prot 数据库，找出灵长目动物(Primates)中所有 alpha 珠蛋白、找出黑猩猩（Chimpanzee）中所有珠蛋白、找出与 HBA_HUMAN 相同位点高于 95%（Identity>95%）的序列。

1. 灵长目所有 alpha 珠蛋白：

Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession	Select for PSI blast	Used to build PSSM	New add
RecName: Full-Hemoglobin subunit alpha; AltName: Full-Alpha-globin; AltName: Full-Hemo...	Homo sapiens	256	256	100%	6e-93	100.00%	142	P00905.2	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	
RecName: Full-Hemoglobin subunit alpha; AltName: Full-Alpha-globin; AltName: Full-Hemo...	Pongo pygmaea	254	254	100%	4e-92	97.89%	142	P06635.2	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	
RecName: Full-Hemoglobin subunit alpha; AltName: Full-Alpha-globin; AltName: Full-Hemo...	Gorilla gorilla	253	253	99%	9e-92	99.29%	141	P01923.1	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	
RecName: Full-Hemoglobin subunit alpha-1; AltName: Full-Alpha-1-globin; AltName: Full-H...	Hylobates lar	253	253	100%	2e-91	98.59%	142	Q8TSS3.2	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	
RecName: Full-Hemoglobin subunit alpha; AltName: Full-Alpha-globin; AltName: Full-Hemo...	Macaca fascicularis	252	252	100%	3e-91	97.18%	142	P53107.2	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	
RecName: Full-Hemoglobin subunit alpha; AltName: Full-Alpha-globin; AltName: Full-Hemo...	Semnopithecus	252	252	99%	4e-91	97.87%	141	P01924.1	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	
RecName: Full-Hemoglobin subunit alpha; AltName: Full-Alpha-globin; AltName: Full-Hemo...	Ateles geoffroyi	252	252	100%	5e-91	96.48%	142	P07817.2	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	
RecName: Full-Hemoglobin subunit alpha; AltName: Full-Alpha-globin; AltName: Full-Hemo...	Sapaeus apella	251	251	100%	7e-91	95.77%	142	P01928.2	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	
RecName: Full-Hemoglobin subunit alpha; AltName: Full-Alpha-globin; AltName: Full-Hemo...	Chlorocebus a...	251	251	100%	7e-91	96.48%	142	P01926.2	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	
RecName: Full-Hemoglobin subunit alpha; AltName: Full-Alpha-globin; AltName: Full-Hemo...	Leontideus f...	251	251	99%	2e-90	96.45%	141	P01929.1	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	
RecName: Full-Hemoglobin subunit alpha-2; AltName: Full-Alpha-2-globin; AltName: Full-H...	Hylobates lar	250	250	100%	2e-90	96.48%	142	Q8TSS4.2	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	

2. 黑猩猩珠蛋白：

更改 Organism 为 chimpanzee

Filter Results

Organism only top 20 will appear ☐ exclude

chimpanzee (taxid:9598)

+ Add organism

Percent Identity to E value to Query Coverage to

PSI-BLAST incl. threshold

0.005

Filter Reset

结果：

Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession	Select for PSI blast	Used to build PSSM	New add
RecName: Full-Hemoglobin subunit alpha; AltName: Full-Alpha-globin; AltName: Full-Hemo...	Pan troglodytes	256	256	100%	6e-93	100.00%	142	P00905.2	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	
RecName: Full-Hemoglobin subunit alpha-3; AltName: Full-Alpha-3-globin; AltName: Full-H...	Pan troglodytes	241	241	99%	1e-86	88.65%	141	P01935.1	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	
RecName: Full-Hemoglobin subunit zeta; AltName: Full-Hemoglobin zeta chain; AltName: Full...	Pan troglodytes	175	175	100%	8e-60	59.15%	142	P06347.2	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	
RecName: Full-Hemoglobin subunit gamma-1; AltName: Full-Gamma-1-globin; AltName: Full...	Pan troglodytes	117	117	98%	6e-36	40.41%	147	P01920.2	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	
RecName: Full-Hemoglobin subunit gamma-2; AltName: Full-Gamma-2-globin; AltName: Full...	Pan troglodytes	117	117	98%	6e-36	40.41%	147	P01921.2	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	
RecName: Full-Hemoglobin subunit beta; AltName: Full-Beta-globin; AltName: Full-Hemo...	Pan troglodytes	117	117	98%	6e-36	43.15%	147	P08873.2	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	
RecName: Full-Hemoglobin subunit delta; AltName: Full-Delta-globin; AltName: Full-Hemo...	Pan troglodytes	115	115	98%	4e-35	43.84%	147	P01772.2	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	
RecName: Full-Hemoglobin subunit epsilon; AltName: Full-Epsilon-globin; AltName: Full-H...	Pan troglodytes	110	110	98%	3e-33	36.99%	147	Q8L3H1.3	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	
RecName: Full-Myoglobin (Pan troglodytes)	Pan troglodytes	61.4	61.4	100%	3e-13	27.52%	154	P02145.2	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	

3.Identity>95%的序列

在 Filter Results 中的 Percent Identity 更改为 “95 to ” 筛选出一个序列。

Filter Results

Organism only top 20 will appear ☐ exclude
chimpanzee (taxid:9598)
[+ Add organism](#)

Percent Identity: 95 to E value: to Query Coverage: to

PSI-BLAST incl. threshold: 0.005

[Filter](#) [Reset](#)

结果:

PSI-BLAST iteration 1													
<input checked="" type="checkbox"/>	Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession	Select for PSI blast	Used to build PSSM	New added	
<input checked="" type="checkbox"/>	RecName: Full=Hemoglobin subunit alpha; AltName: Full=Alpha-globin; AltName: Full=Hemoglob... Pan troglodytes		256	256	100%	6e-93	100.00%	142	P69907.2	<input checked="" type="checkbox"/>			

6.1.2 linux 上的 BLAST 使用

1. .ncbirc

cp /rd1/home/public/blast/.ncbirc . (在根目录下建立文件.ncbirc 进行数据库配置)

.ncbirc 文件内容:

Specifies the path where BLAST databases are installed

[BLAST]

BLASTDB=/rd1/home/blastdb

2.blastp 命令行使用

搜索序列 人α珠蛋白 搜索数据库UniProt Swiss-Prot子库

blastp -query HBA_HUMAN.FASTA -db uniprot_sprot | less

逐屏显示 输出结果

```
Database: uniprot_sprot.fasta
564,638 sequences; 203,519,613 total letters

Query= HBA_HUMAN - P69905, Hemoglobin subunit alpha, HBA1; J Luo,
2016-08-21
Length=142

Sequences producing significant alignments:

      Score  E
    (Bits)  Value
sp|P69907|HBA_PANTR Hemoglobin subunit alpha OS=Pan troglodytes... 286 2e-100
sp|P69906|HBA_PANPA Hemoglobin subunit alpha OS=Pan paniscus OX... 286 2e-100
sp|P69905|HBA_HUMAN Hemoglobin subunit alpha OS=Homo sapiens OX... 286 2e-100
sp|P01922|HBA_GORGO Hemoglobin subunit alpha OS=Gorilla gorilla... 282 1e-98
sp|Q9TS35|HBA1_HYLLA Hemoglobin subunit alpha-1 OS=Hylobates la... 281 3e-98
sp|P06635|HBA_PONPY Hemoglobin subunit alpha OS=Pongo pygmaeus ... 281 4e-98
sp|P01924|HBA_SEMEN Hemoglobin subunit alpha OS=Semnopithecus e... 278 3e-97
sp|P63108|HBA_MACNU Hemoglobin subunit alpha OS=Macaca mulatta ... 278 4e-97
sp|P63107|HBA_MACFU Hemoglobin subunit alpha OS=Macaca fuscata ... 278 4e-97
sp|P67817|HBA_ATEGE Hemoglobin subunit alpha OS=Atelotes geoffroy... 277 1e-96
sp|Q9TS34|HBA2_HYLLA Hemoglobin subunit alpha-2 OS=Hylobates la... 277 1e-96
sp|P18972|HBA_MICAD Hemoglobin subunit alpha OS=Mico argentatus... 276 2e-96
sp|P01926|HBA_CHLAE Hemoglobin subunit alpha OS=Chlorocebus aet... 276 4e-96
sp|P21767|HBA_MACFA Hemoglobin subunit alpha-A/Q/R/T OS=Macaca ... 275 5e-96
sp|P01928|HBA_SAPAP Hemoglobin subunit alpha OS=Sapajus apella ... 275 5e-96
sp|P67818|HBA_SAGOE Hemoglobin subunit alpha OS=Saguinus oedipu... 275 1e-95
sp|P21766|HBA_MACAS Hemoglobin subunit alpha-1/2/3 OS=Macaca es... 275 1e-95
```

搜索序列
α珠蛋白
搜索sw库中
所有人的蛋白质
滤除期望值高于
0.1的假阳性结果
按表格方式在
屏幕输出结果

`astp -query HBA_HUMAN.FASTA -db sp_human -evalue 0.1 -outfmt 6`

行结果如下

- value 指定 E 值过滤假阳性（默认为 10，可选用 0.1 等更小的值）
- outfmt
 - outfmt 6 只输出条目构成的表格，不显示详细比对情况
 - outfmt 7 输出表格加一些表头信息
- word_size 设置种子序列字长
- matrix 设置积分矩（如 PAM250）

迭代PSI-BLAST

构建序列谱计分矩阵时校正参数选择

- 0: 不做校正
1: 基于组分统计校正 (Composition-based statistics, NAR 20:2994-3005)
2: 基于序列特征校正 (Conditioned on sequence properties, Bioinformatics, 21:902-911)
3: 非条件校正 (Composition-based score adjustment, unconditionally, Bioinformatics, 21:902-911)
默认: 2

结果:

最后一行即为第二次搜索得到序列

4.从 fasta 文件建立可用于检索的数据库文件（包括索引）,得到一个自定义搜索库

`makeblastdb -dbtype<nucl/prot> -in input_file [-input_type fasta] -out dbname`

6.2 本小组综合课题选题方向

目前倾向于做植物开花关键调控因子的分子演化。

7. 问题：无

8. 建议：