



序列数据库搜索系统BLAST

—农科院研究生实用生物信息技术课讲稿

2006年10月15日

罗静初

北京大学生物信息中心

Centre of Bioinformatics

Peking University

luojc@pku.edu.cn

<http://www.cbi.pku.edu.cn/>





提 纲

- 何为BLAST?
- 为何BLAST?
- 何处BLAST?
- 如何BLAST?



何为BLAST

- 原文：**Basic Local Alignment Search Tool**
- 直译：基本局部排比搜索工具
- 意译：基于局部序列排比的常用数据库搜索工具
- 含义：蛋白质和核酸序列数据库搜索软件系统及相关数据库
- 用法：以一个或几个蛋白质或核酸序列为检测序列，搜索蛋白质或核酸序列数据库，寻找与检测序列中一个或多个片段具有较高相似性的一组序列，



Search BLAST with Google

(14 Oct 2006)

- BLAST + database: 18,600,000 hits
 - 1st hit: [BLAST tutorial](#)
- BLAST + bioinformatics: 3,310,000 hits
 - 1st hit: [NCBI Tools for Bioinformatics Research](#)
- BLAST + sequence: 14,600,000 hits
 - 1st hit: [BLAST Information](#)



为何BLAST

- 使用方便、功能齐全
- 速度快、结果可信
- NCBI精心维护、持续开发
- 配套数据库不断更新
- 免费服务 (NCBI、EBI、TIGR)
- 免费下载，本地安装



何处BLAST

- NCBI - National Center for Biotechnology Information (US)
- EBI - European Bioinformatics Institute (EU)
- TIGR - The Genome Institute (US)
- Sanger - Sanger Institute (UK)
- UK-CropNet - The UK Crop Plant Bioinformatics Network (UK)
- WU-BLAST - Washington University (US)

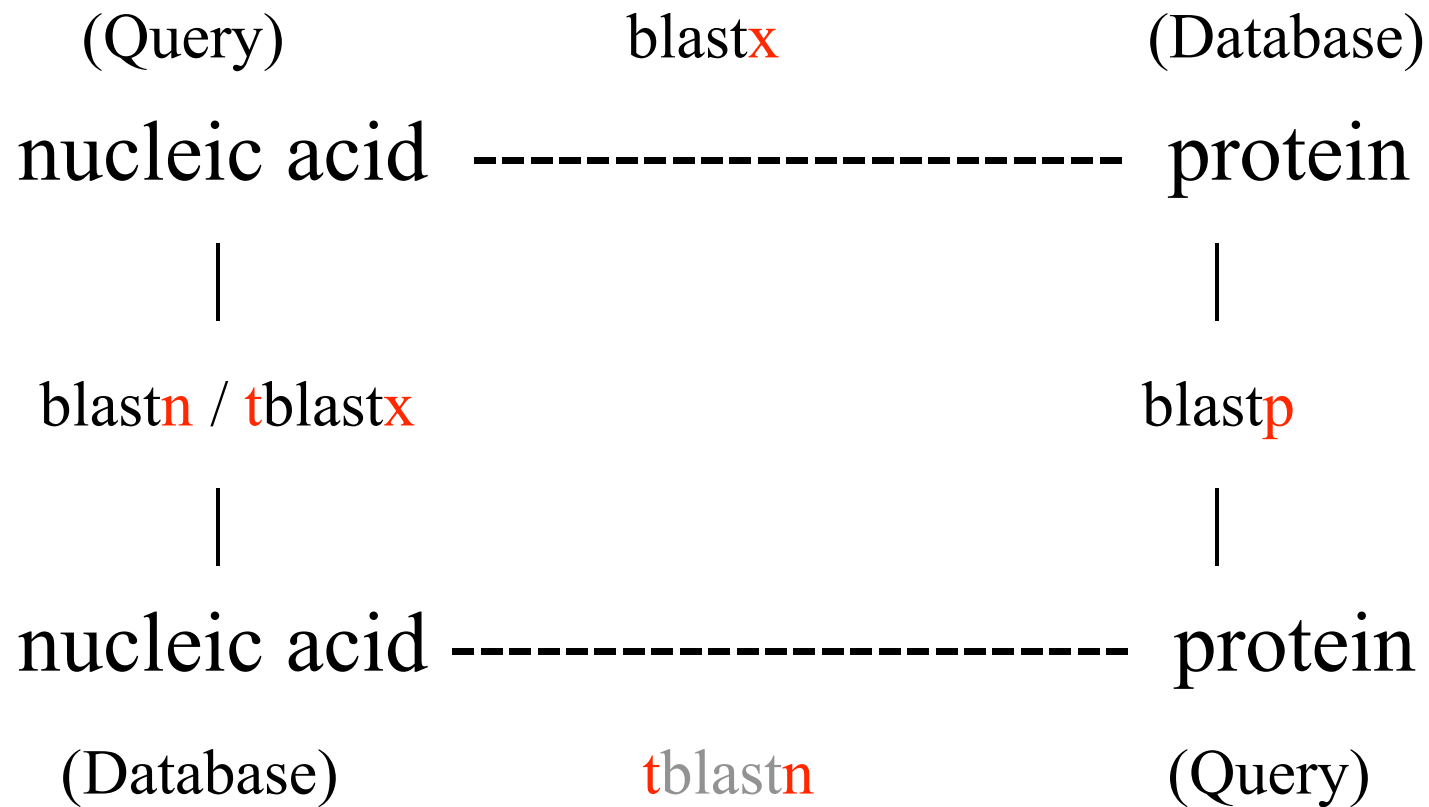


BLAST程序选择

程序	检测序列	数据库
BlastP	蛋白质	蛋白质
BlastN	核酸	核酸
BlastX	核酸	蛋白质
tBlastN	蛋白质	核酸
tBlastX	核酸	核酸



BLAST程序选择





如何BLAST

- Prepare your query sequence
- Select the right database
- Adjust the parameters
- Check the BLAST output
- Make a summary



本地BLAST

- Install the BLAST package
- Create a database in FASTA format
 - Find a set of entries in EXPAsy or SRS
 - Download database from NCBI
 - Construct your own database
- Make BLAST format database
`formatdb -i <database name>`
- Run BLAST locally
`Blastall -p <program> -d <database> -i query
-o <output>`
- Check BLAST output



NCBI常用 BLAST数据库

- **Nucleotide**
 - Patent, EST, GSS, HTG, Chromosome, RefSeq, Non-redundant
- **Protein**
 - Month, Patent, env_pr, PDB, SwissProt, RefSeq, Non-redundant
- **Genome**
 - Human, mouse, rat, chimp, cow, pig, dog, sheep, cat
 - Chicken, puffer fish, zebrafish
 - Fly, honey bee, other insects
 - Microbes, environmental samples
 - Plants, nematodes
 - Fungi, protozoa, other eukaryotes



The great men behind BLAST

- 1970 - Needleman & Wunsch, global alignment
- 1978 - Dayhoff et al., PAM scoring matrix
- 1981 - Smith & Waterman, local algorithm
- 1985 - Lipman & Pearson, database search
- 1987 - Lipman & Pearson, FASTA
- 1990 - Altschul & Lipman, BLAST
- 1992 - Steven Henikoff & Jorja G Henikoff, BLOSUM scoring matrix
- 1997 - Altschul & Lipman, PSI BLAST



可选氨基酸替换计分矩阵

- EMBOSS
 - BLOSUM30, 35, ..., 50, 62, 65, 70, ..., 100
 - PAM10, 20, 30, ..., 100, 110, ..., 500
- EBI WU-BLASTP
 - BLOSUM30, 35, ..., 50, 62, 65, 70, ..., 100
 - PAM10, 20, ..., 100, 110, ..., 500
- EBI FASTA
 - BLOSUM50, 62, 80
 - PAM10, 20, 40, 120, 250
- EBI MPSrch
 - PAM10, 20, ..., 100, 110, ..., 500
- NCBI BLASTP
 - BLOSUM62, 80, 45
 - PAM30, 70



常用计分矩阵比较

	PAM10	PAM250	PAM500	BLOSUM90	BLOSUM62	BLOSUM30
C	10	12	22	9	9	17
S	7	2	1	5	4	4
T	8	3	1	6	5	5
P	8	6	4	8	7	11
A	7	2	1	5	4	4
G	7	5	4	6	6	8
N	9	2	1	7	6	8
D	8	4	3	7	6	9
E	8	4	3	6	5	6
Q	9	4	2	7	5	8
H	10	6	4	8	8	14
R	9	6	5	6	5	8
K	7	5	4	6	5	4
M	12	6	4	7	5	6
I	9	5	3	5	4	6
L	7	6	7	5	4	4
V	8	4	3	5	4	5
F	9	9	13	7	8	10
Y	10	10	15	8	7	9
W	13	17	34	11	11	20