NCBI 简介

ChenYuan

课程概要

- NCBI是啥？有啥？What
- NCBI怎么用？How

  BLAST使用

  BioPython使用

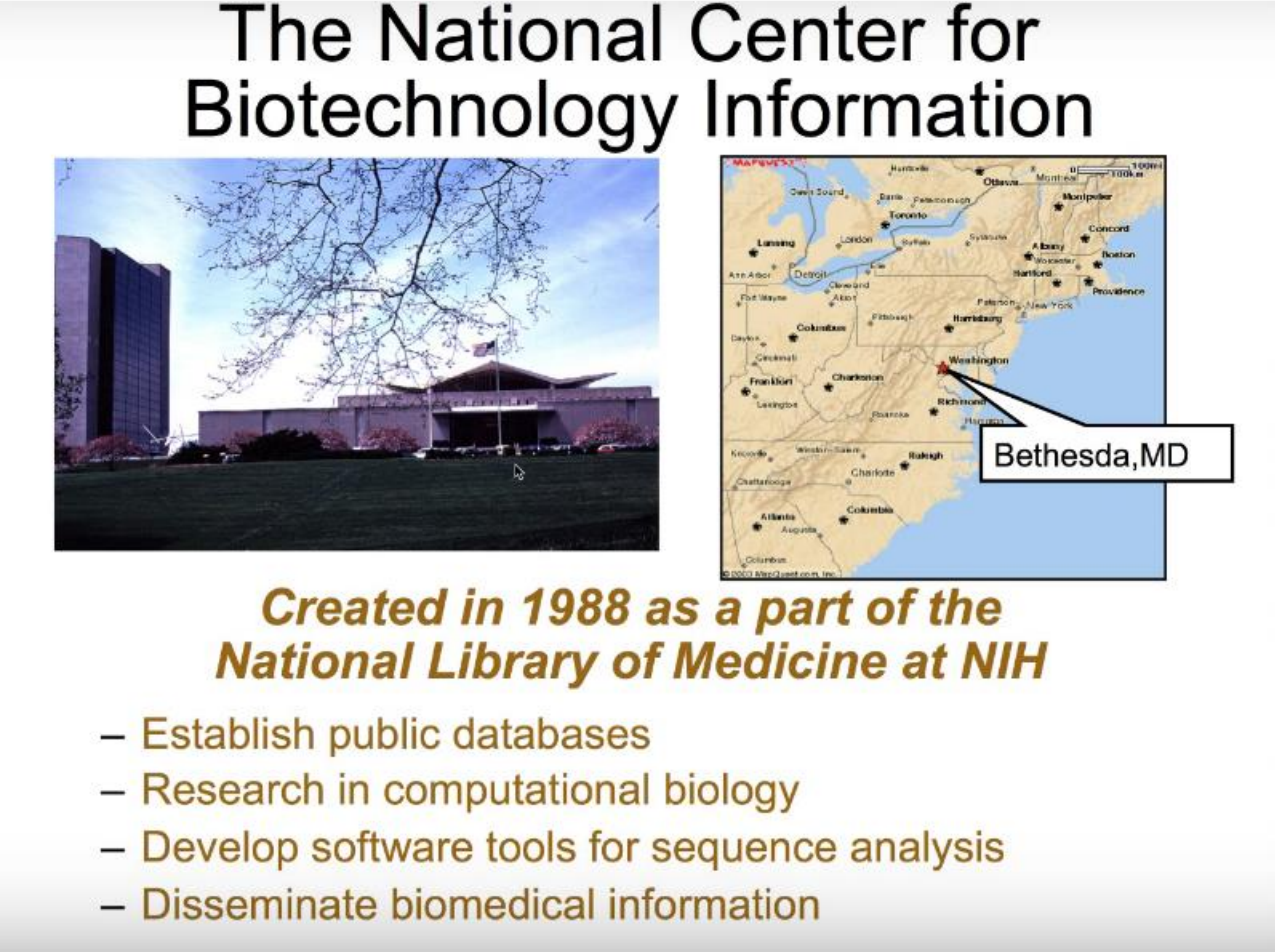
Introduction to NCBI

NCBI是啥？

Introduction to NCBI----What?

国家生物技术信息中心（National Center for Biotechnology Information，简称NCBI）是美国国家医学图书馆（NLM）的一部分（该图书馆是美国国家卫生研究所的一部分）。

Introduction to NCBI----What?



The National Center for Biotechnology Information

Created in 1988 as a part of the National Library of Medicine at NIH

- Establish public databases
- Research in computational biology
- Develop software tools for sequence analysis
- Disseminate biomedical information

Introduction to NCBI----What?

NCBI Databases and Services

- GenBank primary sequence database

- Free public access to biomedical literature
  - PubMed free Medline (3 million searches per day)
  - PubMed Central full text online access

- Entrez integrated molecular and literature databases
- BLAST highest volume sequence search service
  (100 – 200 K searches per day)
- VAST structure similarity searches
- Software and Databases

Introduction to NCBI----What?

NCBI数据库和工具

- **生物数据库**
  - 文献数据库
  - 核酸数据库
  - 基因组资源库
  - ENTREZ数据库
- **生物信息工具**
  - 数据挖掘的工具
  - 序列分析工具
  - 三维结构显示和相似查询的工具
- **图谱（MAPS）**
- **癌症协作研究项目**
- FTP数据下载

NCBI怎么用？

Introduction to NCBI----How?

Entrenz全数据库搜索 Example:血红蛋白β链HBB

查看基因信息

下载FASTA格式数据

Run BLAST

BLAST

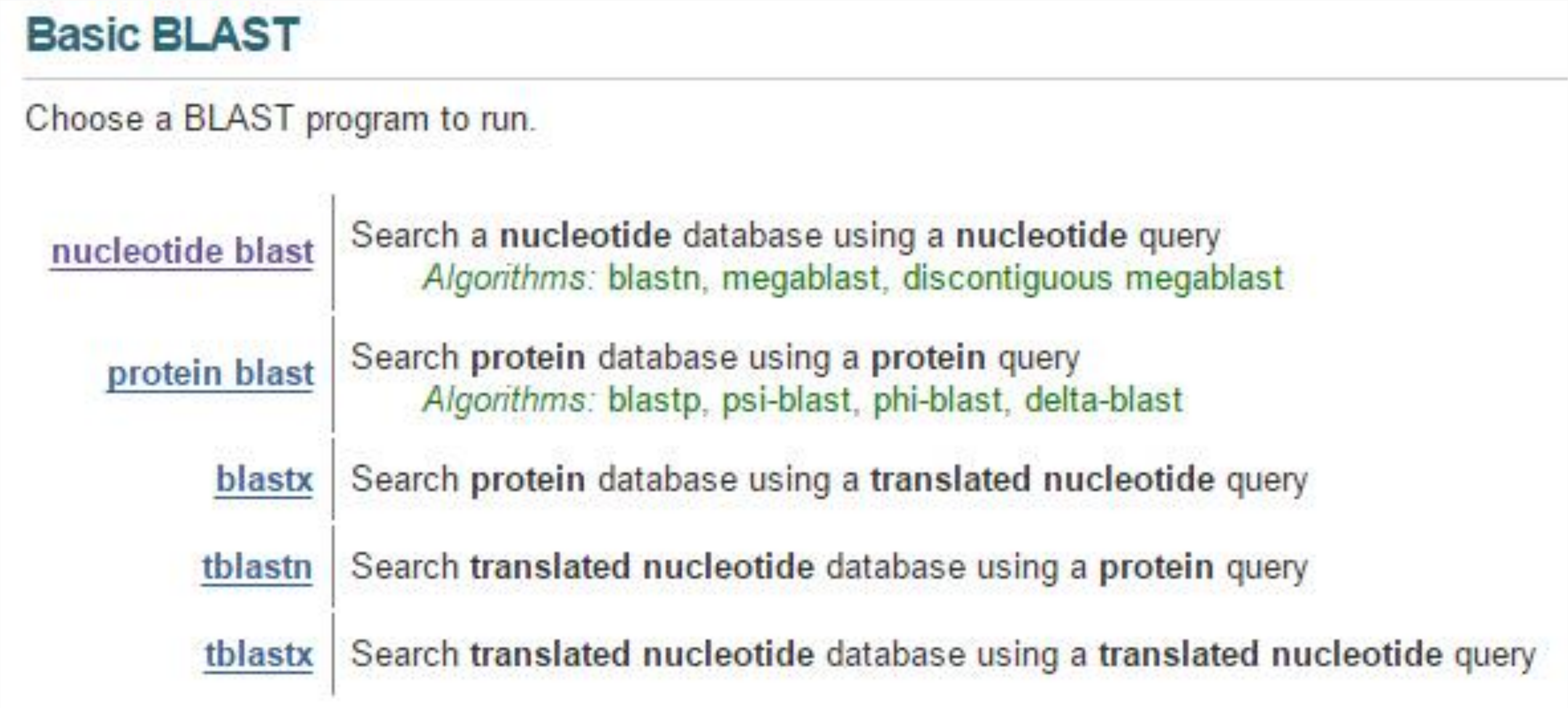
BLAST (Basic Local Alignment Search Tool) allows rapid sequence comparison of a query sequence against a database.

The BLAST algorithm is fast, accurate, and web-accessible.

BLAST--步骤

- 选择一个BLAST程序
- 选择你所感兴趣的序列，输入到BLAST中
- 选择一个用于搜索的数据库
- 为搜索和输出格式选择可选参数

BLAST--Step1 选择一个**BLAST**程序



Basic BLAST

Choose a BLAST program to run.

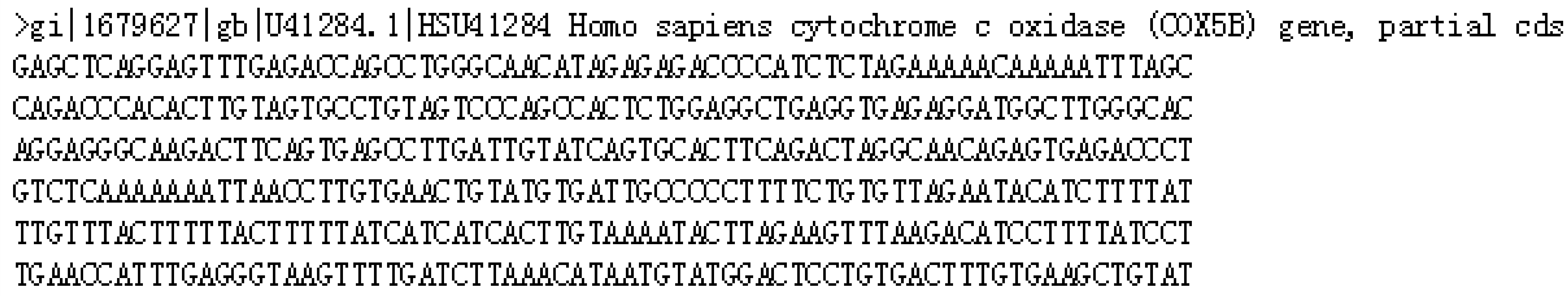
nucleotide blast	Search a **nucleotide** database using a **nucleotide** query *Algorithms:* blastn, megablast, discontiguous megablast
protein blast	Search **protein** database using a **protein** query *Algorithms:* blastp, psi-blast, phi-blast, delta-blast
blastx	Search **protein** database using a **translated nucleotide** query
tblastn	Search **translated nucleotide** database using a **protein** query
tblastx	Search **translated nucleotide** database using a **translated nucleotide** query

BLAST—Step2 输入

可以输入Acession number/gi/FASTA序列
或上传一个文件

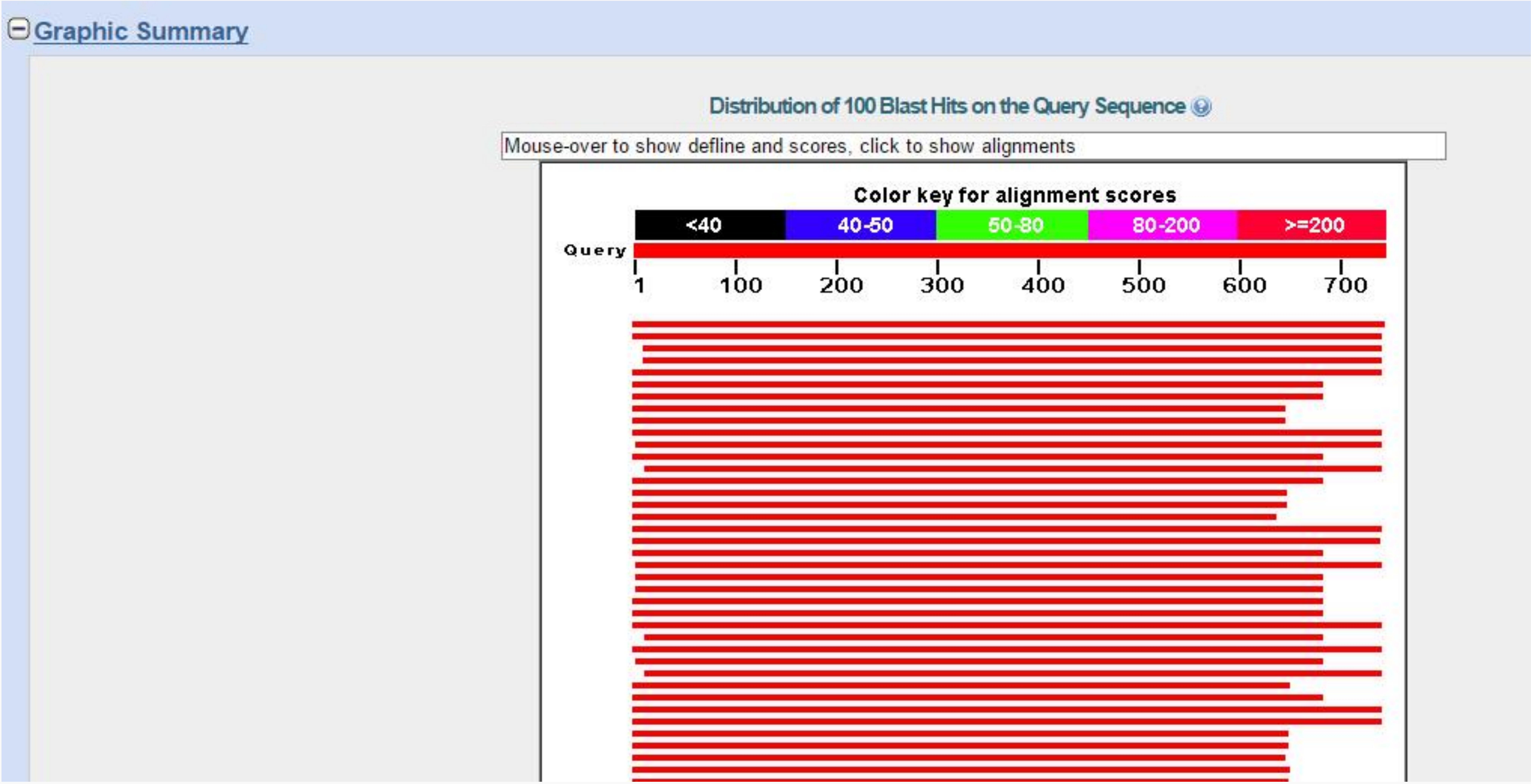
FASTA格式介绍

- 第一行是由大于号“>”起始的任意文字说明
- 从第二行开始为序列本身，只允许使用既定的核苷酸或氨基酸编码符号



```
>gi|1679627|gb|U41284.1|HSU41284 Homo sapiens cytochrome c oxidase (COX5B) gene, partial cds
GAGCTCAGGAGTTTGAGACCAGCCTGGGCAACATAGAGAGACCCCATCTCTAGAAAAACAAAAATTTAGC
CAGACCCACACTTGTAGTGCCTGTAGTCCCAGCCACTCTGGAGGCTGAGGTGAGAGGATGGCTTGGGCAC
AGGAGGGCAAGACTTCAGTGAGCCTTGATTGTATCAGTGCACTTCAGACTAGGCAACAGAGTGAGACCCT
GTCTCAAAAAAATTAACCTTGTGAACTGTATGTGATTGCCCCCTTTTCTGTGTTAGAATACATCTTTTAT
TTGTTTACTTTTTACTTTTTATCATCATCACTTGTAAAATACTTAGAAGTTTAAGACATCCTTTTATCCT
TGAACCATTTGAGGGTAAGTTTTGATCTTAAACATAATGTATGGACTCCTGTGACTTTGTGAAGCTGTAT
```

BLAST—Step3 Result



BioPython

- Download: http://biopython.org/wiki/Download
- Example:
- 用BioPython转录翻译
- 用BioPython进行BLAST