
A Field Guide to GenBank and NCBI Resources

Medha Bhagwat
Peter Cooper
Susan Dombrowski
Andrei Gabrielian
Chuong Huynh
Wayne Matten
Rana Morris
Steve Pechous
Vyvy Pham
Eric Sayers
Tao Tao

bhagwat@ncbi.nlm.nih.gov
cooper@ncbi.nlm.nih.gov
dombrows@ncbi.nlm.nih.gov
gabrieli@ncbi.nlm.nih.gov
huynh@ncbi.nlm.nih.gov
matten@ncbi.nlm.nih.gov
morris@ncbi.nlm.nih.gov
pechous@ncbi.nlm.nih.gov
pham@ncbi.nlm.nih.gov
sayers@ncbi.nlm.nih.gov
tao@ncbi.nlm.nih.gov

updated 08/02/04

Online Resources

Lecture Resources:

Field Guide Home: <http://www.ncbi.nlm.nih.gov/Class/FieldGuide/>
Power Point Slides: <ftp://ftp.ncbi.nih.gov/pub/FieldGuide/Slides/Current/>
Problem Set: ftp://ftp.ncbi.nlm.nih.gov/pub/FieldGuide/NCBI_exercises.pdf
Glossary: <http://www.ncbi.nlm.nih.gov/Class/FieldGuide/glossary.html>

General:

NCBI Homepage: <http://www.ncbi.nlm.nih.gov>
Site Map: <http://www.ncbi.nlm.nih.gov/Sitemap/index.html>
About NCBI: <http://www.ncbi.nlm.nih.gov/About/>
NCBI News: <http://www.ncbi.nlm.nih.gov/About/newsletter.html>
NCBI Handbook: <http://www.ncbi.nlm.nih.gov/books/bv.fcgi?call=bv.View..ShowTOC&rid=handbook.TOC&depth=2>

GenBank:

GenBank Release Notes: <ftp://ftp.ncbi.nih.gov/genbank/gbrel.txt>

Collaborating Nucleotide Databases:

EMBL: <http://www.ebi.ac.uk/>
DDBJ: <http://www.ddbj.nig.ac.jp/>

Entrez:

Entrez: <http://www.ncbi.nlm.nih.gov/Entrez/>

BLAST:

BLAST Main Page: <http://www.ncbi.nlm.nih.gov/BLAST/>

BLAST statistics: <http://www.ncbi.nlm.nih.gov/BLAST/tutorial/Altschul-1.html>

Frequently Asked Questions: http://www.ncbi.nlm.nih.gov/BLAST/blast_FAQs.html

BLAST Guide: <http://www.ncbi.nlm.nih.gov/BLAST/producttable.html>

BLAST Clients, Executables and Databases: <ftp://ftp.ncbi.nih.gov/blast/>

NCBI Source Code: ftp://ftp.ncbi.nih.gov/toolbox/ncbi_tools/

NCBI Structures:

Structure Homepage: <http://www.ncbi.nlm.nih.gov/Structure/>

Cn3D tutorial: <http://www.ncbi.nlm.nih.gov/Structure/CN3D/cn3dtut.html>

CDD Search: <http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi>

CDart: <http://www.ncbi.nlm.nih.gov/Structure/lexington/lexington.cgi?cmd=rps>

Genomic Resources:

Entrez Genomes: <http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=Genome>

Genomic Biology: <http://www.ncbi.nlm.nih.gov/Genomes/>

Human Genome Resources: <http://www.ncbi.nlm.nih.gov/genome/guide/human/>

Map Viewer <http://www.ncbi.nlm.nih.gov/mapview/static/MVstart.html>

LocusLink: <http://www.ncbi.nlm.nih.gov/LocusLink/>

UniGene: <http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=unigene>

Human Genome Sequencing: <http://www.ncbi.nlm.nih.gov/genome/seq/>

Human Genome BLAST: <http://www.ncbi.nlm.nih.gov/genome/seq/HsBlast.html>

Spidey: <http://www.ncbi.nlm.nih.gov/IEB/Research/Ostell/Spidey/>

ePCR (UniSTS): <http://www.ncbi.nlm.nih.gov/genome/sts/epcr.cgi>

Mouse Genome Resources:	http://www.ncbi.nlm.nih.gov/genome/guide/mouse/
Trace Archive Megablast:	http://www.ncbi.nlm.nih.gov/blast/tracemb.html
Mouse Genome BLAST:	http://www.ncbi.nlm.nih.gov/genome/seq/MmBlast.html

Other Databases:

SWISS-PROT	http://us.expasy.org/sprot/
PIR	http://pir.georgetown.edu/pirwww/pirhome.shtml
PDB	http://www.rcsb.org/pdb/
PRF	http://www.prf.or.jp/en/

Contact:

<i>General Help</i>	info@ncbi.nlm.nih.gov
<i>BLAST</i>	blast-help@ncbi.nlm.nih.gov
<i>Help Desk Hotline</i>	301-496-2475

Literature Reference List**General**

Baxevanis, A. and Ouellette, B.F.F., eds. *Bioinformatics: A Practical Guide to the Analysis of Genes and Proteins. Second edition* New York: John Wiley & Sons. 2001. ISBN: 0-471-38391-0

Gibas, C. and Jambeck, P. *Developing Bioinformatics Computer Skills*. Sebastopol: O'Reilly and Associates. 2001. ISBN:1-56592-664-1

Mount, D. W. *Bioinformatics: Sequence and Genome Analysis*. Cold Spring Harbor Laboratory Press. Cold Spring Harbor. New York. 2001. ISBN: 0-87969-608-7

Wheeler DL, et al. 2004. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.* **32**(1):35-40. PMID: 14681353

GenBank Database

Benson DA, et al. 2004. Genbank : update. *Nucleic Acids Res.* **32**(1):23-26 PMID: 14681350

Ouellette BF, Boguski MS.1997. Database divisions and homology search files: a guide for the perplexed. *Genome Res.* **7**(10):952-5. PMID: 9331365

BLAST

Altschul SF, *et al.* 1990. Basic local alignment search tool. *J Mol Biol.* **215**(3):403-10. PMID: 2231712.

Altschul SF, *et al.* 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**(17):3389-402. PMID: 9254694.

Altschul SF, *et al.* 1998. Iterated profile searches with PSI-BLAST--a tool for discovery in protein databases. *Trends Biochem Sci.* **23**(11):444-7. PMID: 9852764.

Schaffer AA, *et al.* 1999. IMPALA: matching a protein sequence against a collection of PSI-BLAST-constructed position-specific score matrices. *Bioinformatics.* **15**(12):1000-11. PMID: 10745990.

Tatusova TA, *et al.* 1999. BLAST 2 Sequences, a new tool for comparing protein and nucleotide sequences. *FEMS Microbiol Lett.* **174**(2):247-50. PMID: 10339815.

Zhang Z, Schwartz S, Wagner L, Miller W. 2000. A greedy algorithm for aligning DNA sequences. *J Comput Biol.* **7**(1-2):203-14. PMID: 10890397.

Zhang Z, *et al.* 1998. Protein sequence similarity searches using patterns as seeds. *Nucleic Acids Res.* **26**(17):3986-90. PMID: 9705509.

MMDB, Cn3D, CDD and Structures

Chen J, *et al.* 2003. MMDB: Entrez's 3D structure database. *Nucleic Acids Res.* **31**(1):474-477. PMID: 12520055.

Marchler-Bauer *et al.* 2003. CDD: a curated Entrez database of conserved domain alignments. *Nucleic Acids Res.* **31**(1): 383-387. PMID: 12520028.

Tatusov RL, *et al.* 2003 The COG database: an updated version includes eukaryote. *BMC Bioinformatics* **4**(1):41. PMID: 12969510

Specialized Genomic Resources

Jang W, *et al.* 1999. Making effective use of human genomic sequence data. *Trends Genet.* **15**(7):284-6. PMID: 10390628.

Pruitt KD *et al.*, 2003 NCBI Reference Sequence Project: update and current status *Nucleic Acids Res.* **31**(1):34-37. PMID: 12519942.

Sherry, S.T., *et al.* 2001. dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res.* **29**(1):308-311. PMID: 11125122

Tatusov RL, *et al.* 2001 The COG database: new developments in the phylogenetic classification of proteins from complete genomes. *Nucleic Acids Res.* **29**(1):22-28. PMID: 11125040