



## Study Guide to

### Database Resources of the National Center for Biotechnology Information

*Nucleic Acids Research, 2007, Vol.35, Database issue D5-D12*

The NCBI home page web site is: \_\_\_\_\_.

The Entrez database includes ~ \_\_\_\_\_ DNA and protein sequences derived from several sources. It also includes over \_\_\_\_\_ gene-oriented sequence clusters in UniGene, \_\_\_\_\_ genetic variations in dbSNP and over \_\_\_\_\_ protein structures from the Molecular Modeling Database. How many online scientific textbooks does the database contain? \_\_\_\_\_

Results retrieved from an Entrez search may be sent to the Entrez \_\_\_\_\_ where they may be recalled later.

### PubMed and PubMed Central

The PubMed database includes over \_\_\_\_\_ citations from greater than \_\_\_\_\_ life science journals for biomedical articles back to the 1950s.

PubMed Central is a digital archive of peer-reviewed journals in the life sciences providing access to greater than \_\_\_\_\_ - \_\_\_\_\_ articles!

Participation in PMC requires a commitment to \_\_\_\_\_ to full text.

The NCBI taxonomy database is growing at the rate of \_\_\_\_\_ new taxa a month.

BLAST means:

\_\_\_\_\_  
(find this answer by looking at the 9<sup>th</sup> reference entry)

BLAST programs perform \_\_\_\_\_ similarity searches returning to the researcher a set of gapped alignments. Each returned alignment is scored and assigned a measure of statistical significance called the Expectation Value (\_\_\_\_\_).

BLAST even takes into account the \_\_\_\_\_ composition of the query sequence in its estimation of statistical significance. A new \_\_\_\_\_ option

creates a dendrogram that clusters sequences according to their distances from the query sequence.

The database of single nucleotide polymorphisms (\_\_\_\_\_) is a repository for single base nucleotide \_\_\_\_\_ and short \_\_\_\_\_ and \_\_\_\_\_ polymorphisms. This database contains over \_\_\_\_\_ human SNPs and another \_\_\_\_\_ from a variety of other organisms.

Within the tools for gene-level analysis, the \_\_\_\_\_ (ORF) performs a six-frame translation of a nucleotide sequences.

An electronic PCR can be performed to estimate the genomic binding site, \_\_\_\_\_ and specificity for sets of primer pairs in specific organisms. This site is [www.ncbi.nlm.nih.gov/sutils/e-pcr](http://www.ncbi.nlm.nih.gov/sutils/e-pcr).

The Gene expression omnibus (\_\_\_\_\_) is a database for \_\_\_\_\_ and other forms of molecular data generated by the scientific community. The repository contains ~\_\_\_\_\_ individual measurements, derived from ~3000 array definitions, and spanning \_\_\_\_\_ organisms.

The Online Mendelian Inheritance in Man (OMIM) provides a catalog of human genes and \_\_\_\_\_. It contains information on disease \_\_\_\_\_ and genes. The database contains ~\_\_\_\_\_ entries.