PBLAST
Go to the NCBI BLAST website:
https://blast.ncbi.nlm.nih.gov/Blast.cgi

1. What are the differences among the following blast programs. What is used as the input ("query") and what type of database is searched?

blastn, blastp, blastx, tblastn, tblastx

2. Try a blastp search of the UniProtKB/Swiss-Prot database with the following query sequence:

YDVEKNNSRIKLGLKSLVSKGTLVQTKGTG

How long did the search take? Can you identify the protein? Is it possible to identify the species from this short peptide sequence? What is the E-value of the best match? How many matches are there with $E < 10^{-10}$?

3. Now try the same search with the following peptide sequence of the same length that was generated at random:

VLAMMPKLRQWCPMKLPRTQSSWAYGHEDT

How long did the search take? What is the result? Are there any "good hits"? What is the E-value of the best match?

4. Do a blastp search of the SwissProt database for the following sequence:

mmqesatetisnssmnqngmstlssqldagsrdgrssgdtssevstvellhlqqqqedvvsytqvic

What protein does it match? From which species? What is the accession number of the human homolog of this protein?

5. Use the accession number of the above protein for a blastp search of the chimpanzee (*Pan troglodytes*) genome using the Reference proteins (refseq_protein) database.

What is the percent identity between the human and the chimp protein? How many amino acid differences are there?

6. Go to the UCSC genome browser at http://genome.ucsc.edu/

and do a BLAT search of the *Drosophila melanogaster* genome for the following two primer sequences. Will these primers work for PCR in *D. melanogaster*? What gene do they match? What is the expected length (in bp) of the PCR product? (You may also try this with the "in silico PCR" tool of UCSC)

5' CGGCTCGCCACTCCAAATTGG   3'

5' GTGCTGCAGCAACTGTTGCTG   3'