Bioinformatics Resources
Exercise Sheet 2
due on May 29th, 9 a.m.
Total number of points: 10 P

Task 1
For the institution you work for, you should write a parser which digests information from PDB from scratch that will be used for the next ten years. Which are three file formats, available to you, which one would you choose and why? (2P)

Task 2
Look at the most represented organisms in PDB. Is that also the organism, the proteins are expressed in for the structure determination? Is the distribution of the source organisms closer to UniProtKB/Swiss-Prot or TrEMBL? Why is that? (2P)

Task 3
What does the resolution of a PDB entry represent and what unit is it typically denoted in? Do all of the three main experimental methods supply a resolution? Which not? Which entry currently has the highest resolution (lowest value) in PDB? What is the average resolution in the PDB? (2P)

Task 4
Write an Perl script that retrieves a plain text PDB file. One or more PDB ids are given as simple arguments. The result is stored as file <PDBID>.pdb.
Example : (2P)

>./getPDB.pl 1c3w 1brd

Task 5
Determine the RMSD (Root-Mean-Squared Deviation) between the structures of 1c3w and 1brd. (2P)

Good luck,
J. Reeb & L. Richter