

CMSC423 Homework 2  
Handed out: 2/7/07  
Due: 2/14/07

This is one of the few assignments that require you to write code in Perl.

Please write a Perl script that can help retrieve data from the NCBI Trace Archive using the *query\_tracedb* script provided by NCBI (see lecture notes for details of where to find this script). For this assignment you will have to perform the following tasks:

1. Download the *query\_tracedb* script and learn how to use it.
2. Write a Perl script that takes on the command line a simple Trace Archive query and retrieves the relevant data from this repository. Note that the NCBI places limits on the number of records that can be retrieved from their servers. For this exercise assume that you are only allowed to retrieve 5,000 records at a time. Thus, your Perl script may have to, transparently to the user, perform multiple calls to *query\_tracedb* in order to retrieve all the data.

Additional details:

1. A sample query, to get you started, is:  
SPECIES\_CODE = "WOLBACHIA ENDOSYMBIONT OF DROSOPHILA  
MELANOGASTER"
2. Please retrieve just the fasta and quality information, in tar.gz format.
3. It is OK if your script generates multiple output files, corresponding to individual chunks of 5,000 records.
4. Please provide us with a simple README file that indicates how to run your program. Make sure you include a sample invocation that you know for sure will work.

**Grading!** We will grade all aspects of the code, including how “pretty” it looks.

Specifically pay attention to the following aspects:

1. Please make sure that your code works as advertised in the README file you provided. If your code doesn’t work as indicated in the README file you will automatically lose 50% of the grade for this assignment.
2. Please provide copious comments and format your code so that it is easy to read. Part of your grade will be based on the formatting of the code.

Please contact me as soon as possible if you have any questions regarding this assignment, or if you “get stuck” and might not be able to complete the assignment on time. Once the assignment is due I will no longer accept any excuses.

**Good Luck!**