Hands-on perl

Before starting, download and install a text editor that support Perl colors:

For mac user, TextWrangler:
http://www.barebones.com/products/textwrangler/

For Linux user, Emacs:
http://ftp.gnu.org/pub/gnu/emacs/

If you are a Windows user and would like to do this exercise on your computer, you first need to install Perl. The easiest way is to install ActivePerl. You should also install a text editor that supports Perl colors: http://sourceforge.net/projects/emeraldeditor/

In this exercise, you will try to modify the blast parser that we have just described so that your parsed file include a new column that corresponds to the "Identities" values.

1. Create a working directory that you will name perl_exercise (remember yesterday's primer on Linux). The location of this directory is not important but a suggestion is in your home (on a mac: /Users/username). To move to your home: cd ~

2. Copy in this directory the blast output (blast_output.txt) and the Perl script that you will modify (blast_parsing.pl)

3. Open the blast output file in the text editor that you have just installed and identify the lines that contain what we are looking for. Here is one such line:

   ```
   Identities = 80/200 (40%), Positives = 111/200 (55%), Gaps = 21/200 (10%)
   ```

4. With the structure of these lines in mind (pay attention to every details), think about a regular expression that would recognize them and pick only the identity values (ex. 80/200 (40%)). You can use this website for help with designing your regular expression: http://regexpal.com/

   Think about a way to print only the values corresponding to the best hits.

5. Open and modify in the text editor the Perl script blast_parser.pl (keep the original version, and save the file that you are modifying as blast_parser_exercise.pl)

6. To test your new script, type on the command line:

   ./blast_parser_exercise.pl blast_output.txt > blast_parsed.txt

7. Don't forget to chmod a+x your script before the first try, and remember that with Perl "there is more than one way to do it"