



4/14/2008





Opening Discussion

- Any questions about anything?



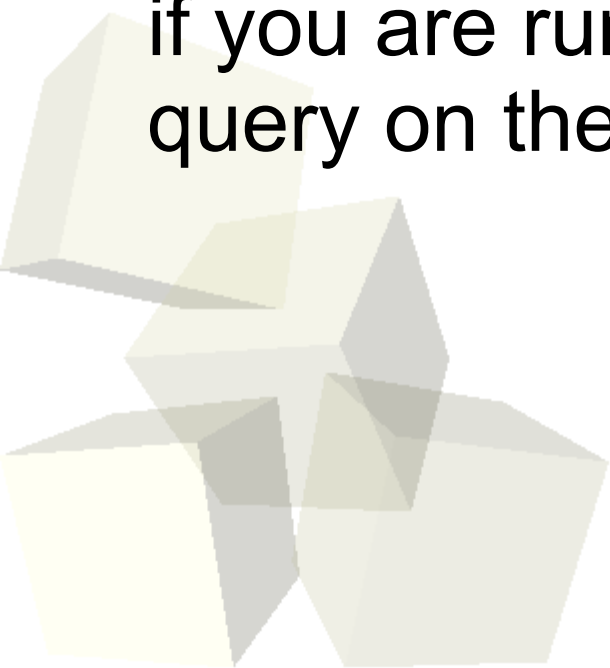


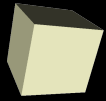
File Test Operators

- Last class I mentioned how you can open folders and read the names of the files in them.
- Perl has lots of operations to get information about those files. They all have the form of a hyphen followed by a character.
- Examples include
 - ◆ -e tells if a file exists
 - ◆ -f tells if the file is a plain file
 - ◆ -d tells if the file is a directory
 - ◆ -s gives the size of the file in bytes
- Pages 323 and 324 have a list of these.



- The Basic Local Alignment Search Tool (BLAST) is a program that allows you to compare sections of DNA to one another.
- You can run BLAST over the web or download a local copy and run it on your machine. Which one you pick depends on your needs. Running Perl scripts on the output probably makes more sense if you are running locally, but we can go do a query on the web to see what we get as output.





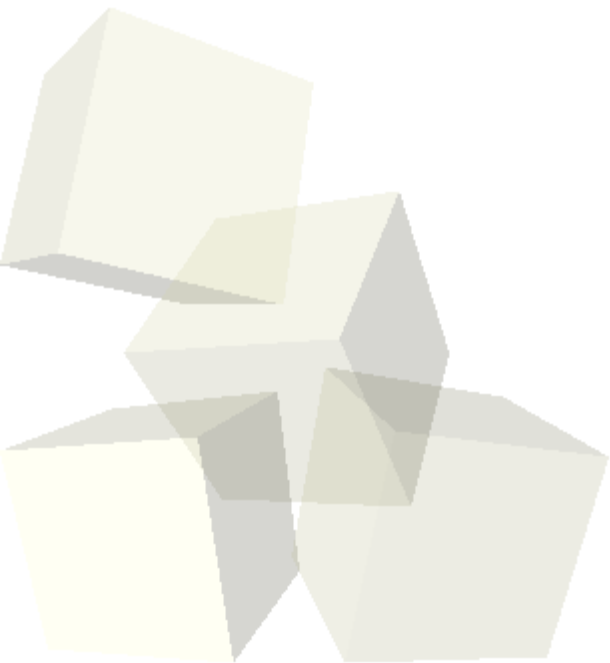
Parsing BLAST Output

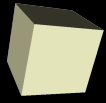
- The BLAST output has three basic sections. There are beginning and ending annotations with the alignment in between.
- The alignments section begins with “ALIGNMENTS” and the end annotation begins with “ Database:” so we can split things out with a regular expression.
- This chapter shows the use of assigning a match on a regular expression directly into a list. The list elements become \$1, \$2, ...



New RegEx Features

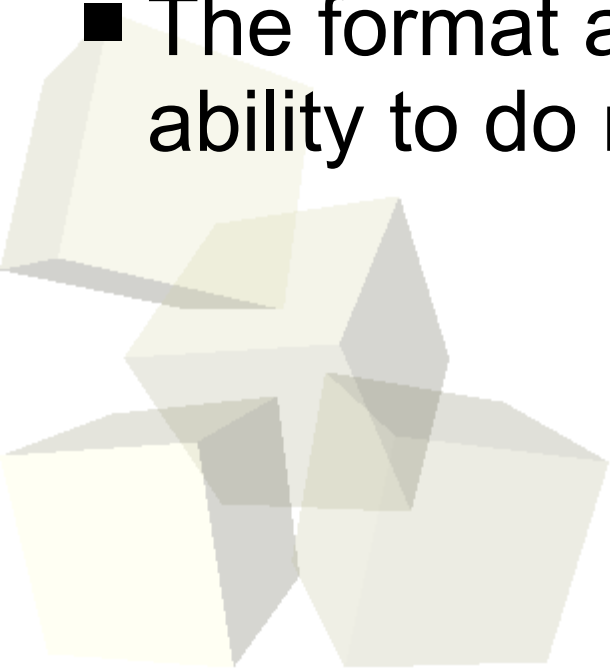
- To parse the alignments the chapter introduces some other features as well.
- Global matches can be put inside a while loop to loop through all matches in the string.
- `(?!>)` is a negative lookahead assertion.
- `*?` and `+?` are non-greedy repetition matches.





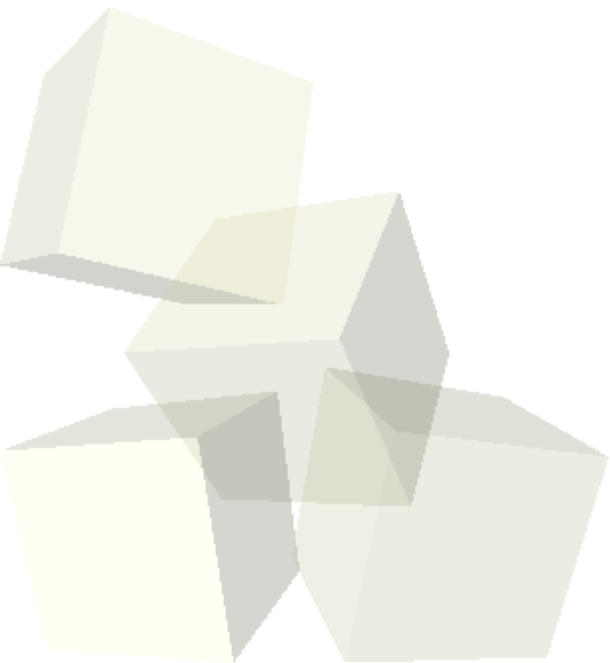
Data Presentation

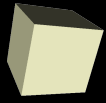
- Perl includes other features that help us print out data in nicely formatted ways.
- The printf function works very much like the version in C.
- here documents are good for printing out multiple lines of output. (Though I find them a bit odd to read in in the code.)
- The format and write commands give you the ability to do report style printing.





- The last section on chapter 12 gives an introduction to the Bioperl package. This is a module for Perl that has been written to assist people with doing bioinformatics. If you spend much time using Perl for these types of tasks you probably want to familiarize yourself with this package.





Closing Comments

- Assignment #8 is due today.
- Quiz #5 will be next class.
- The first student presentation is this Friday. Who wants it?

