Presentation:

* Discuss NCBIFetch and its final configuration (TIFFANY)

In the execution of our accession number the user would input before the program runs its G4 sequence search, I found NCBI Fetch as the best method to start this process. Instead of having a personal repository to our interface, NCBI Fetch is a means to upload sequences by a GenBank accession number or a GenBank GI (NCBIFetch, 2008). The objects and method provided by the BioJava CookBook modified a starting point sequence due to the implementation of the Biojavax extension, which created difficulty in past methods to obtaining sequences from NCBI (NCBIFetch, 2008). In the initial run of the NCBI Fetch with our main class, the whole GenBank file displayed as our output, therefore I parsed out the sequence and am working on a means to write the sequence into a FASTA file. For week 9, along with Sarah Woodruff, we are to write the sequence into a FASTA file, then have it run successfully in our main class.

NCBIFEtch (2008). BioJava: CookBook: external sources: NCBIFetch. Retrieved from <http://biojava.org/wiki/BioJava:CookBook:ExternalSources:NCBIFetch>.

BioJava is an [open-source](http://en.wikipedia.org/wiki/Open_source) project dedicated to providing a [Java](http://www.java.sun.com/) framework for processing biological data. It provides analytical and statistical routines, parsers for common file formats and allows the manipulation of sequences and 3D structures. The goal of the biojava project is to facilitate rapid application development for bioinformatics.

http://biojava.org/wiki/Main\_Page

* Final version of regex search (initial to final version to show evolution) (TIFFANY)

Another major goal is in the algorithm design. The backbone of our algorithm uses a BioJava utility, Regex (regular expression search), which was incorporated to serve as our pattern matcher in our program. The default algorithm we used is (as stated earlier) (G>/=3Nx G>/=3Nx G>/=3Nx G>/=3Nx), but with one to seven bases in between the three or more guanine repeats (G3+N1-7 G3+N1-7 G3+N1-7 G3+N1-7). Within our program the regex pattern was a followed: ([gG]{3,}\\w{1,7}){3,}[gG]{3,}, where [gG] describe the upper or lower case of guanine found in the genome , {3,} describe a parameter set of 3 or more guanines, \ quotes the following character, \w denotes a word character ( a nucleotide character), and {1,7} describing a parameter set between one and seven. The parameter set for one to seven nucleotides between the triplet or more guanines found only 35 G4s within the 2127 bp chromosomal genome. After trial and error from Lindsay Cook and Tiffany Younger, a new mode of expanding the limit of found nucleotide characters between one and thirty, and including an overlap between each match pattern, generated 976 found G4 motifs.