

Due: - Written Exercises due Wednesday, April 27<sup>th</sup>, at 4 pm in Homework box in 2131 Kemper;  
 - Program due electronically Wednesday, April 27<sup>th</sup>, at 11:50 pm.

Written Exercises (25 pts): The written exercises should be typed and each page should have at the top your name and ID#, section, and hw#. Handwritten answers will not be graded.

J&K, 5.6.4, 5.6.8, 5.6.12, 6.2.10, 6.2.16, 6.4.4, 6.4.6, 6.4.16, 6.5.2, 6.6.2, 6.6.4.

Program (75 pts):

What to handin: a makefile and source code files with the name specified, and have your makefile produce the executable file specified (use **a11** ... at the top). The third line in the source code files must contain the author of the file, ID, and section #. Use the *handin* program for electronic submission, described in the UNIX tutorial. For this homework use:

```
handin cs30 hw4 Makefile <your file1> <your file2> ...
```

The date and time your files are created in the cs30 directory will be counted as your submit times. If those times are later than 11:50 pm on the due date your submissions will be considered late.

### DNA Palindromes, forward and reverse sequences

Technological advances nowadays allow scientists to read the DNA sequence of any organism, and find the genes in those sequences by using computational in addition to biological and chemical methods. The area of computer science that deals with such problems is called **bioinformatics**. In this homework you will be asked to analyze input DNA sequences and find patterns in them.

The genetic code of all living organisms is written in their DNA, which chemically is a double helix molecule. Abstractly, DNA is a double string over the characters **a**, **c**, **g**, **t** (representing four nucleic acids) where for each character **a**, **c**, **g**, and **t** on the first string there lies a complementary character **t**, **g**, **c**, and **a** on the second string, respectively. Thus, **a** is complementary to **t** and **c** to **g**. The two complementary strings are called strands. For example the following is a short DNA sequence:

```
aacgtttgcgcgat
|||||
ttgcaaacgcgcta
```

Many properties of DNA molecules are encoded as patterns of letters in their sequences thereby allowing us to compare molecular similarity and deduce molecular evolution in organisms. One important pattern in DNA is whether a sequence is palindromic, i.e. if it reads the same forward and back. E.g. **aaacgttgcaaa** is a palindrome.

Write a program that asks the user for an input DNA sequence, in the form of a single strand (ie a single string), repeatedly until 'q' is entered. Check if the sequence is a proper DNA sequence (ie consists of only the letters A, a, C, c, T, t, G, g) and then process the sequence and produce the following output. If the sequence is a palindrome print out "This sequence is a palindrome." If it isn't print out that it isn't. Then, print the input sequence and its reverse complement (ie the second strand, read back to front).

Make sure the output of your program matches exactly the output below of my executable located at `/home/cs30/public/hw4/dnapal` on the csif machines. In the same directory you will find a file **help.c** which contains a function declared as `void strrev(char *, int);` which reverses a given string (you don't need to use it, but it may be helpful).

[cs30@pc50 hw4]\$ codons

Input DNA string (use only a, A, c, C, g, G, t, T) or 'q' to exit:

aaatttaacgatatagcaagcag

---

- (1) This sequence is not a palindrome.
- (2) Forward sequence: AAATTTAACGATATAGCAAGCAG.
- (3) Reverse complementary sequence: CTGCTTGCTATATCGTTAAATTT.

Input DNA string (use only a, A, c, C, g, G, t, T) or 'q' to exit:

atggggaaaaagagaccctctagcataa

---

- (1) This sequence is not a palindrome.
- (2) Forward sequence: ATGGGGAAAAAGAGACCCTCTAGCATAA.
- (3) Reverse complementary sequence: TTATGCTAGAGGGTCTCTTTTTCCCCAT.

Input DNA string (use only a, A, c, C, g, G, t, T) or 'q' to exit: ata

---

- (1) This sequence is a palindrome.
- (2) Forward sequence: ATA.
- (3) Reverse complementary sequence: TAT.

Input DNA string (use only a, A, c, C, g, G, t, T) or 'q' to exit: ATGTGA

---

- (1) This sequence is not a palindrome.
- (2) Forward sequence: ATGTGA.
- (3) Reverse complementary sequence: TCACAT.

Input DNA string (use only a, A, c, C, g, G, t, T) or 'q' to exit: qwerty

---

Input DNA string (use only a, A, c, C, g, G, t, T) or 'q' to exit: q

[cs30@pc50 hw4]\$