

CS250 Assignment 1 DNA p-distance Matrix

Date assigned: Friday, February 1, 2013
Date due: Friday, February 8, 2013
Points: 30

The purpose of this assignment is to produce the p-distance matrix for a file of DNA strings.

Consider two DNA strings s_1 and s_2 of equal length. The p-distance, denoted $p(s_1, s_2)$, is the proportion of corresponding symbols that differ between s_1 and s_2 .

Given the following DNA strings:

```
AAAACCCGGT
AAACCCGGGT
AACCCCGGGG
AAAACCCCGG
```

The p-Distance matrix is:

```
0.0 0.2 0.4 0.2
0.2 0.0 0.2 0.4
0.4 0.2 0.0 0.4
0.2 0.4 0.4 0.0
```

You are to write a C++ program that will read a collection of at most 10 equal length (at most 80 bases) DNA strings from a file called dnastrings.txt. Your output is to look exactly like the following:

```
*****
DNA p-distance matrix
*****
```

DNA Strings

```
AAAACCCGGT
AAACCCGGGT
AACCCCGGGG
AAAACCCCGG
```

p-distance Matrix

```
0.0 0.2 0.4 0.2
0.2 0.0 0.2 0.4
0.4 0.2 0.0 0.4
0.2 0.4 0.4 0.0
```

Goals for Assignment 1

1. Implement a C++ program that uses files, 1D & 2D arrays, and character processing
2. Break up a program into well-defined functions
3. Test your program one function at a time
4. Implement a basic program used in Bioinformatics

Notes:

1. Your main function is to be mainly variable declarations and function calls.

To complete this assignment you must submit the following:

1. An electronic copy of your program on Turing
 - a) Create a new C++ solution in Visual Studio 2010. Your solution must be called **PUNetIDAssignments**. For example, mine would be called ryandjAssignments. This solution will hold all of your assignment projects. You are starting over for CS250, meaning do not include any projects from CS150.
 - b) Your project for this assignment must be named **01_DNA**. It is vital that you name your solution and your project correctly!
 - c) Type your program (fully documented) into the project. You need to follow the coding standards from the CS250 Web page. These coding standards have been modified to include additional C++ language features introduced in CS250, so please be sure to read the new coding standards.
 - d) Pay attention to the example output! Your program's output must look exactly like the example output! The spacing and newlines in your output must match exactly.
 - e) Make sure that your program compiles without errors & warnings and runs correctly. If you get any errors or warnings, double check that you typed everything correctly. Be aware that C++ is case-sensitive. You will lose 10% if there are any warnings and 70% if your program does not successfully compile.
 - f) Once you are sure that the program works, it is time to submit your program. You do this by logging on to Turing and placing your complete solution folder in the proper Drop folder.
 - g) The solution must be in the drop folder by time class starts on the day the assignment is due. Anything submitted after that will be considered late.
2. A hard copy of your program
 - a) The hard copy must be placed on the instructor's desk by the time class starts on the day that it is due.
 - b) The hard copy must be printed in color, double-sided, and stapled in the upper left corner if your solution contains multiple pages. I do not bring a stapler to class.

Remember, if you have any problems, come to me straight away with your project on a flash drive or on Turing. Good Luck!!!! ☺