

CS250 Assignment 1

DNA p-distance Matrix

Date assigned: Friday, January 30, 2015

Date due: Monday, February 9, 2015

Points: 25

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

The p-distance between two DNA strings (s_1 , s_2) is the proportion of corresponding bases that differ between the two strings.

For example, if

- S_1 is: AAAACCCGGT
- S_2 is: AAACCCGGGT

Then the p-distance, denoted $p(s_1, s_2)$ is $2/10$ or 0.2 . The differing bases are the 4th and 7th ones.

Write a complete C++ program that will read in a collection of at most 20 equal length (at most 80 bases) DNA strings from a file called dnastrings.txt. You are then to create a p-distance matrix for those strings.

Input: The file named (dnastrings.txt) contains an unknown number of DNA strings. There will be at most 20 strings and the length of each string will not exceed 80 bases. Each string will be separated by a new line.

```
AAAACCCGGT
AAACCCGGGT
AACCCCGGGG
AAAACCCCGG
```

Your **output** is to look exactly like the following:

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

DNA Strings

```
S1 : AAAACCCGGT
S2 : AAACCCGGGT
S3 : AACCCCGGGG
S4 : AAAACCCCGG
```

p-Distance Matrix

	S1	S2	S3	S4
S1	0.0	0.2	0.4	0.2
S2	0.2	0.0	0.2	0.4
S3	0.4	0.2	0.0	0.4
S4	0.2	0.4	0.4	0.0

Goals for Assignment 1

1. Reacquaint yourself with Visual Studio and Grace.
2. Implement a C++ program that uses files, 1D & 2D arrays, and character processing.
3. Break up a program into well-defined functions. This is the first assignment where you choose the functions. It is important that your functions be small, and focused on a particular task.

Notes:

1. Your main function is to be mainly variable declarations and function calls.
2. Test your program one function at a time.

To complete this assignment you must submit the following:

1. An electronic copy of your program on Grace

- a) Create a new C++ solution in Visual Studio 2010. Your solution must be called **PUNetID-Assignments**. For example, mine would be called khoj0332-Assignments. This solution will hold all of your assignment projects for CS 250. 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/commented) 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 sample output. The spacing and newlines in your output must match exactly.
- e) Make sure that your program builds 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 build successfully.
- f) Once you are sure that the program works, it is time to submit your program. You do this by logging on to Grace and placing your complete solution folder in the **CS250-01 Drop** folder.
- g) The solution must be in the drop folder by the 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.
- c) Your tab size must be set to 2 and you must not go past column 80 in your output.

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