

DNA Extra Credit problem (10 hw points)

In class, we spent time solving the problem of shotgun sequencing of DNA. For this process, you need to put together a long string from many small ones; the way we do this is finding overlap between the end of one string and the beginning of another, continuing this process until we've put together as much as possible. This is a complicated problem, so we started by making some simplifying assumptions; we "undid" these assumptions, except the following:

- There are exactly two strings
- We will only compare the end of dna1 to the beginning of dna2.

For this problem, you will "undo" the remaining assumptions, and finish a complete C++ program that does shotgun sequencing. The only built-in C++ function you may use is length. NOTE: you will not necessarily be able to come up with a single string from the array.

Our partial pseudocode solution was:

```
overlap = 10
match = False
while overlap < length(dna1) and overlap < length(dna2) and match == False
    match = True
    j = len(dna1) - overlap // index into dna1
    k = 0 // index into dna2
    while match == True && k < overlap:
        if (dna1[j] != dna2[k]):
            match = False
        j = j + 1
        k = k + 1
    overlap++
```

Follow this link: <http://math.scu.edu/~linnell/cs10w16/EC-DNAarray.txt> for an array of strings you can use to practice with.

Please note: This problem is HARD. It is worth 10 EC points, but will probably take you longer than a (whole) regular homework assignment.