Genome Assembly - Greedy Algorithm

Authors: Tony Kabilan Okeke, Ifeanyi Osuchukwu
Date: 02.24.2022

Write a function `fastq_assemble_greedy( fastqfile_OR_reads )` in a separate `fastq_assemble_greedy.m` (or `fastq_assemble_greedy.py`) file, that takes in either the name of a fastq file containing short reads, or a cell array of short reads (a list of short reads in Python), implements the "greedy algorithm" of genome assembly, and returns the assembled contigs as a cell array of texts (in python, return a list of texts).

The greedy sequence assemble algorithm includes the following steps:

- Find two reads having the longest overlap;
  - merge these two reads into a single, longer read.
  - If there are more than two pairs with the same longest overlap, use the one that would produce a merged sequence that comes before the other alphabetically.
- Repeat the previous step so long as at least two reads can be found sharing one or more residues overlap.
- Return the collection of extended, merged reads you end up with.

Once you complete the `fastq_assemble_greedy()` function, run the test cases below and save this report as a pdf file, where the output of these cases are shown. You do not need to make any changes in any of the codes below.

In [ ]:
```
# Imports
from fastq_assemble_greedy import fastq_assemble_greedy
import bmes
```

Test Case 1

In [ ]:
```
fastq_assemble_greedy( ['AAA','AAB','ABB','BBA','BBB'] )
```

Out[ ]:
```
['AAABBA', 'BBB']
```

Test Case 2

In [ ]:
```
fastq_assemble_greedy( ['AAA','AAAA','AAAAA','BBB','BBBB','BBBBB'] )
```

Out[ ]:
```
['AAAAA', 'BBBBB']
```

Test Case 3

In [ ]:
```
file = bmes.downloadurl("http://sacan.biomed.drexel.edu/lib/exe/fetch.php?rev=&media=course:binf:genomeassembly:hwgenome
fastq_assemble_greedy( file )
```

Out[ ]:
```
['AAABBA', 'BBB']
```

Test Case 4

In [ ]:
```
file = bmes.downloadurl("http://sacan.biomed.drexel.edu/lib/exe/fetch.php?rev=&media=course:binf:genomeassembly:hwgenome
fastq_assemble_greedy( file )
```

Out[ ]:
```
['FGHIABCDEFGHIABCDGH']
```

Test Case 5

In [ ]:
```
file = bmes.downloadurl("http://sacan.biomed.drexel.edu/lib/exe/fetch.php?rev=&media=course:binf:genomeassembly:hwgenome
fastq_assemble_greedy( file )
```

Out[ ]:
```
['AAABBA', 'BBB', 'CCC']
```