

Computational Genomics: Sequences

Homework 1

Srinivas Suresh

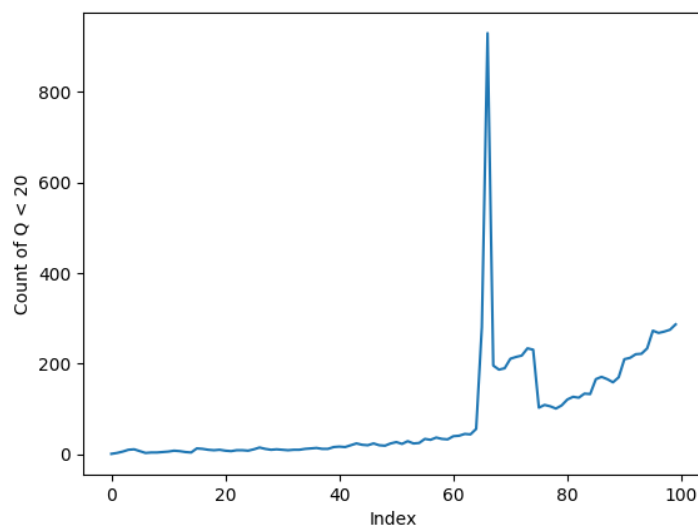
September 29, 2017

Answers

Some answers are instructions on how to run the associated program

1. `answer1.py > inputfile < outputfile`

- (a) Yes, the sequencer could not determine the 67th nucleotide quite often in every read



(b)

There appears to be an increase in bases with $Q < 20$ as the index increases. This tells me that the longer the sequence the lower the quality of bases at the end

2. Winner indicated in **Bold**
 - (a) Bad Character Rule: 2 moves, 1 skip
Good Suffix Rule: 8 moves, 7 skips
 - (b) **Bad Character Rule: 8 moves, 7 skips**
Good Suffix Rule: 4 moves, 3 skips
 - (c) Bad Character Rule: 5 moves, 4 skip
Good Suffix Rule: 10 moves, 9 skips
3. Winner indicated in **Bold**, the answer for case b changes
 - (a) Bad Character Rule: 2 moves, 1 skip
Good Suffix Rule: 8 moves, 7 skips
 - (b) **Bad Character Rule: 8 moves, 7 skips**
Good Suffix Rule: 8 moves, 7 skips
 - (c) Bad Character Rule: 2 moves, 1 skips
Good Suffix Rule: 10 moves, 9 skips
4. Send a pattern in via stdin, ensure complete works file is in same dir is in the same dir To run, call *answer4.py <pattern_file*

The data for the provided patterns was as follows

Word	Exact	1-Mismatch	Specificity
achievements	3	3	0.0582
acquaintance	36	0	0.2608
remembrances	5	64	0.2584

5. Send a pattern in via stdin, ensure complete works file is in same dir is in the same dir To run, call *answer5.py <pattern_file*

The data for the provided patterns was as follows

Word	Exact	1-Mismatch	Specificity
achievements	3	3	0.667
acquaintance	36	0	0.5
remembrances	5	64	0.9324