

Project

Project

- Problem: Given M number of short reads of length L , reconstruct the original sequence of length N that those short reads come from.
 - one page mid report – team, data(how you are going to make or get data), WHAT (define your problem), HOW(idea), WHEN(timeline)
 - Construct an algorithm and implement it. On the week of presentation, you should check your implementation with TA. You are going to download your program from eclass that you uploaded before the presentation and use it.
 - Presentation.
 - Final report. You should upload and “PRINTOUT” your report before the class start at the day of your presentation. If you don’t prepare the printout, you cannot present.

Things you can consider

- Complications that you could consider
 - N could be long (up to 3,000,000,000)
 - M could be large (typically 20million~ 200 million)
 - Length of L (typically 32~100)
 - number of mismatches in a read, D
 - There could be repeats (sequence repeats in several places)
 - single-end or paired – end reads
 - denovo or reference? SNP location?
 - else – find by yourself, just define well in the presentation.

Presentation

- presentation - 4~5 min, question- 1 min. Two questions for other presentations
- Must include the following slides
 - Introduction
 - Clearly define your problem
 - Explain your data (ex. where you got the data, how you generate the data, or etc.)
 - Input and Output
 - Benchmark - Other algorithm that you compared with your algorithm
 - Your Algorithm
 - Result
 - about your machine. (unix/mac/pc, CPU, memory size, etc)
 - time and space complexity in either big O notation or actual time/space
 - Compare with the benchmark
 - Future work
 - cons and pros of your algorithm
 - What you can do to improve your algorithm

Evaluation and some tips

- Evaluation
 - Difficulty – how difficult your problem is.
 - Accuracy – algorithm and implementation accurately did what is should do.
 - Clearness – clear presentation, two reports
 - Two questions- this is MUST
- Tips
 - Try to be different from others.
 - Same presentation, better present early.
 - Finish implementation as least a three days before the presentation. you need time to practice presentation. “CEARL PRESENTATION IS IMPORTANT”

Something you should “NOT”

- TIME. you cannot present longer than 5 mins. Practice presentation!!!
- Never show code. You can run program if necessary but never show codes. Explain your algorithm not code.
- Never copy and paste papers. This is not a paper review presentation. Please don't put too much WORDS or SENTENCES.
- you need OUTPUT– if the problem is hard you may fail to get good result. But AT LEAST you need results to show.