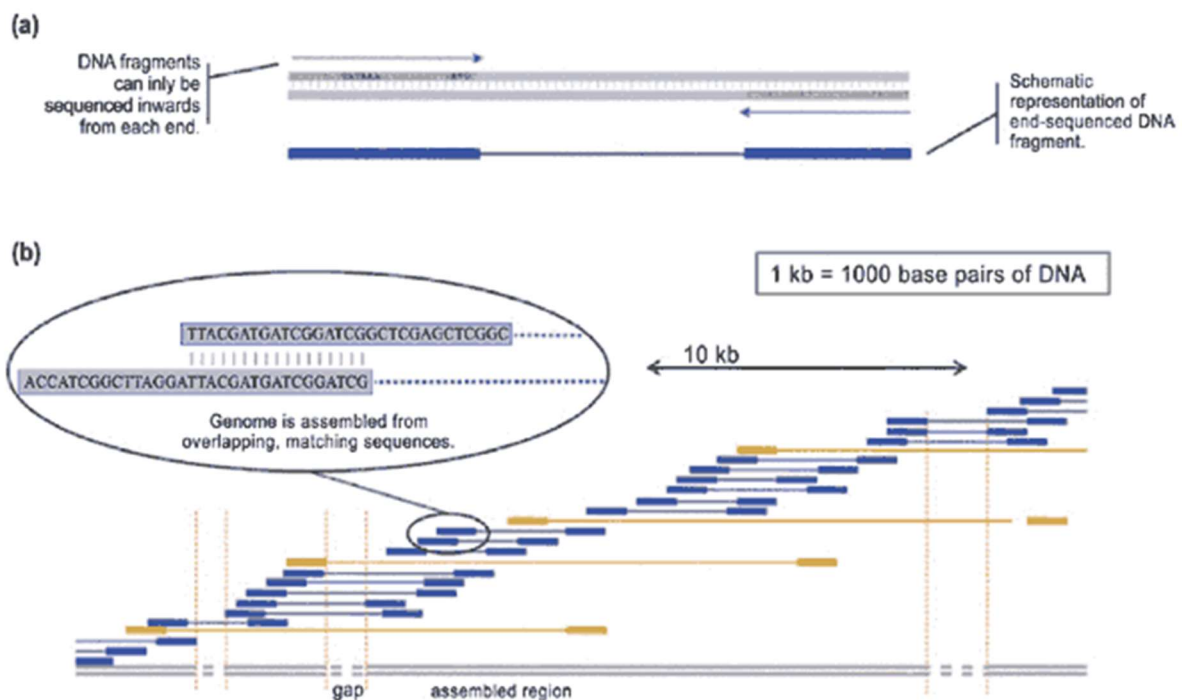


Abstract:

My project domain will be in genomics. I am a studying genetics and biochemistry and eventually I would like to go into bioinformatics, which is the use of computer science and math to organize and analyze biological (usually genetic) data.

Part I:

I will be exploring bowtie, a short read alignment tool. When DNA is sequenced, it is broken up into many little pieces so that its sequence of bases (ATGC) can be recorded in parallel. There are different ways to do this, but no matter what you end up with short lists of pair sequences called reads. A computer has to put them back together in order in a process called alignment. Bowtie is one such tool, and, because genetic data is large, it has to be parallelized.



My goal for part of one of the project is to identify strategies in the code to parallelize it and set up an example with instructions.

Part II:

My goal for part two is to take what I learned from part one and write my own parallelized algorithm capable of aligning a small subset of short reads, noting the strategies I used to speed it up and benchmarking where possible. I don't yet know to what degree of success I can aspire. However, I want to understand how this works, and there's no better way than by doing. I would consider this operation to be a success if I can get this working and demonstrably speed it up.

Schedule:

- Week of February 5: C++ Refresher (Project Proposal Milestone **Due 2/5**)
- Week of February 12: Reading Code
- Week of February 19: Reading More Code & Writing
- Week of February 26: Rest (Project Part 1 **Due 2/26**)
- Week of March 5: Algorithm Psuedo Code & Sample Identification
- Week of March 13: Get it working
- Week of March 19: Get it working
- Week of March 26: Probably Still Trying to Get it working & Parallelizing
- Week of April 2: Parallelizing
- Week of April 9: Writing Report (Final Project due **4/15**)