Project Update: Mutant HIV Sequence Difference Visualization

Michael DiBernardo
November 16th, 2005

Motivation

- HIV is composed of a pair of complementary RNA strands
- That is, we can generate one sequence given the other
- So, we can think of HIV as a character sequence over the alphabet {A, C, G, U}

Virology in a nutshell

- When a virus infects a host, it mutates
    - Different strains can arise in the same organism

Virology in a nutshell

- These changes can manifest themselves as:
  - Substitutions (SNPs)
  - Insertions
  - Deletions
- All with respect to a single ‘canonical’ HIV viral sequence

So what?

- People in Africa have been found that are immune to HIV
  - They carry the disease but are not affected by it

So what?

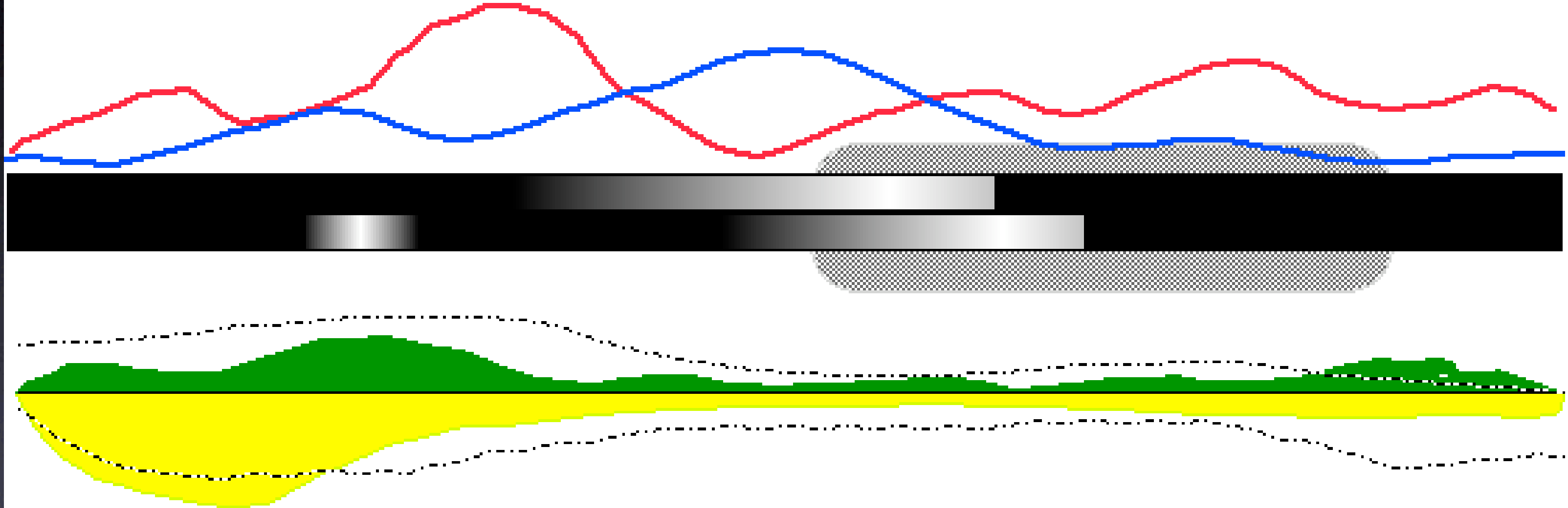
- What if the virii in the unaffected population are undergoing some unusual changes?

The experiment

- Dr. Steven Jones and Dan Baluta have isolated thousands of viral mutants from immune and susceptible populations
- They are currently computing the differences between these mutants and the canonical viral sequence

My role

- They want two things:
  - A single static visualization that will immediately communicate important results for a publication
  - An explorative tool



Progress

- Discussed requirements with clients
- (Attempted) to draft some task scenarios with them
- Decided on data communication format
- Designed overview

Progress

- Played with (lots of) existing codebases
- Decided on Java, Swing, and JFreeChart (originally was going to use openGL)
- Implemented data model

Challenges

- No real data (yet)
- Clients are very difficult to get a hold of
- Main user is located in Winnipeg
- Clients have a very narrow focus -- do not even want to think about interaction

Suggestions

- If you have any suggestions or criticisms, I would be extremely happy to hear them