Nick Egan

Anna Ritz

BIO 131 - Intro to Computational Biology

4/18/17 (late)

Project Proposal - Utilizing GraphSpace to Showcase DNA Differentiation

In this project, I hope to utilize GraphSpace in a way that showcases the various ways that DNA differentiates: deletions, insertions, inversions, and translocations. Methods of graphically displaying these methods certainly exist, as in the form of dot plots, and simple diagonal line graphs. However, I intend to use GraphSpace as a way of potentially adapting these styles of comparison to significantly larger datasets, while maintaining graphical simplicity. In doing so, I intend to adapt the tool of GraphSpace to determine its usability for comparing between datasets, and finding effective graphical methods of showing those differences.

This will be done by determining a method to computationally find at least deletions, insertions, and hopefully inversions and translocations as well. Then, the tools that exist in GraphSpace will be explored to see what may be effective for translating complicated comparative information into simple graphics - for example using methods of coloring connections, compressing non-differentiated sections of DNA into single nodes for simplicity, and the like. This will be tested with smaller, artificial datasets at first to make sure that methods of catching structural rearrangements works, and then this will be used to create graphs. The quality of these graphs will be scored by readability, and how much information is lost in their creation, which are fairly qualitative measures. Hopefully by the end, these methods will be adaptable to significantly larger, real datasets, such as those from the GenBank, BioProject, or other NCBI sources of genetic information.

The two hardest parts of the project will likely be, first, making a working system that can find structural rearrangements, and second, figuring out how to use GraphSpace in an effective way. I have hopes that this project will turn out with at least the first part surpassed, and working on the second section will provide a greater foundation for my upcoming summer research in the Biology department.