

DMP Final Report: Development of a Genome Visualization Tool

Kristi Blumenberg
Mentor: Dr. Cindy Grimm

I. Introduction

Gathering gene data from different genomes is essential to understanding individual genes' functions and evolutionary history. However, genomes with thousands or millions of genes make it difficult to compare them as a whole, understand how genes in the genome are functionally dispersed, locate gene clusters, or predict operons. Visualization tools are useful for this purpose by generating an image given a set of genes. The purpose of this project is to develop a tool to efficiently read Escherichia Coli and Bacteroides Thetaiotaomicron gene data and draw circular genome maps, including operon predictions and gene cluster information.

II. Previous Work

Visualization tools exist with a wide variety of functions. Some only draw one map and display information for one specie, others can do this for many different species. The amount of detailed information also varies greatly depending on the use of the tool. I aimed to produce a tool similar to E. Coli K12 Zoom Tool [1] and the KEGG E. Coli K12 tool [2], because of the straightforward user interface, searching features, and zooming features. With both tools, the user selects an area of the genome map to zoom in on, which is displayed in another window. The user can then use arrow buttons to move to other positions in the genome.

On the extreme side, the UCSC Genome Browser [3] is an incredible tool that compares genome data of many animal species and displays detailed information of the genome at different locations. Once my tool is fully functional, I will use this tool to determine what other types of information and data could be read in and displayed in the map.

III. Data and Methods

The E. Coli and B. Theta data were gathered through the research of Professor Jeremy Buhler of Washington University. The first objective of this project was to determine the best way of zooming in to the map to provide a linear view of the genes at a certain location and their positions and lengths relative to each other. I studied the research of Sheelagh Carpendale to learn more about magnification in context with fish-eye warping, which essentially pushes a portion of the image forward toward the viewer while the rest of the image remains the same. Using C++ and OpenGL, I created a square mesh and texture mapped the image created

after reading the data onto the mesh. Using FLTK, I implemented several features for the user interface:

- Gene information: gene, location, PID, strand, code, length, and cluster of orthologous groups (COG)
- Search for a gene given its name
- Selecting a genome
- Move the warp clockwise or counter-clockwise
- Increase/Decrease zoom
- Increase/Decrease warp radius to view more genes

See Figure 1.

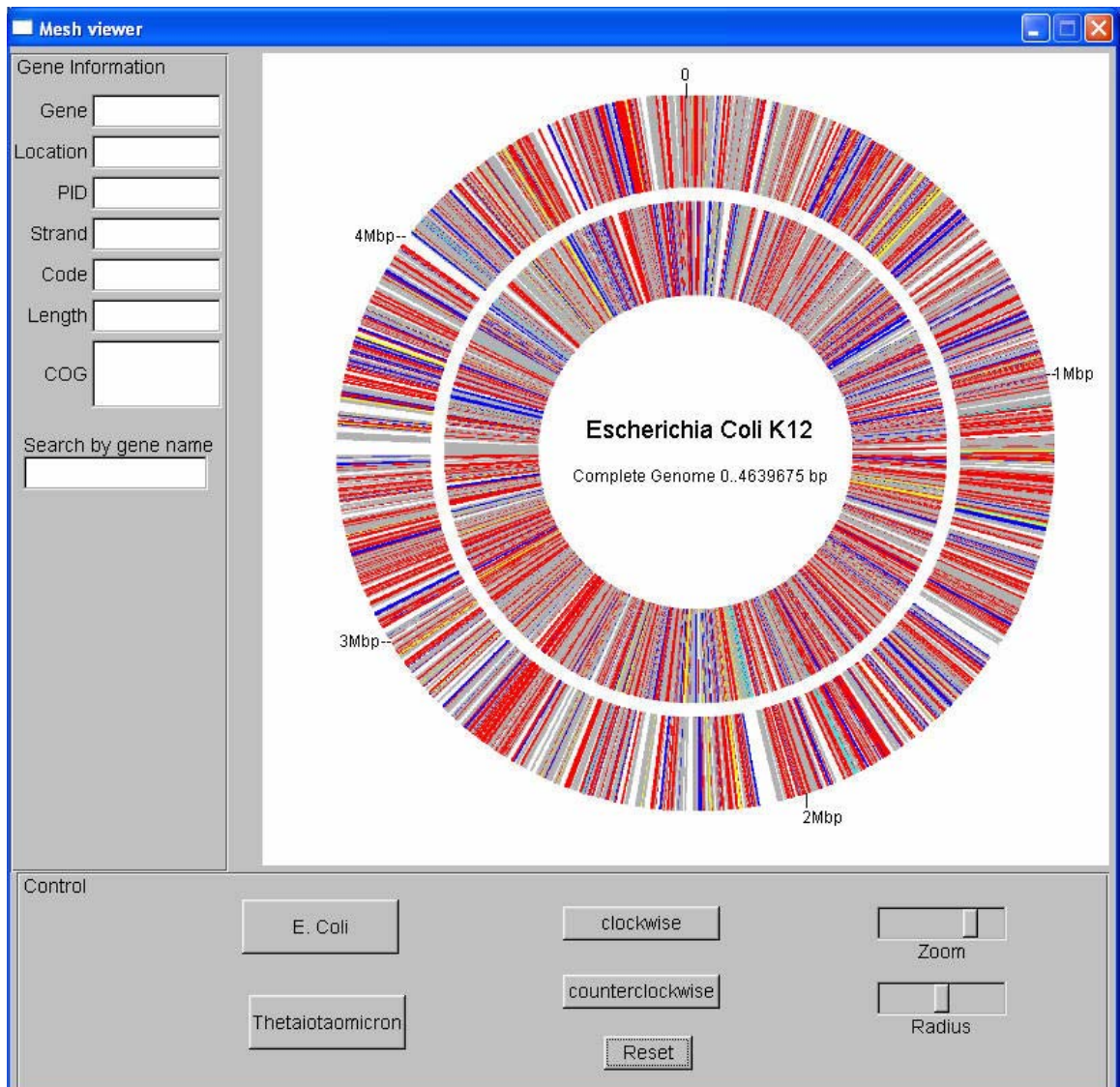


Figure 1: Genome Map User Interface

However, the fish-eye warping did not to increase the resolution of the selected area of the genome enough to view the specific genes. Instead, I created a

circular mesh and produced a rectangular texture map, which proved to be the more efficient method for creating the image because there were much fewer calculations. For zooming, if a face of the mesh was selected as the zooming face, its scale would be increased and magnified. To eliminate overlapping of faces, the rest of the faces in the mesh were scaled smaller so that the magnified face would fit into its proper location in the map. Having the zoomed face in the same position is a unique method of zooming, because most other tools use a separate window where its location in the whole genome is unknown. When the user wanted to view adjacent genes, the clockwise and counter-clockwise buttons could be used to select an adjacent face to magnify. Increasing the zoom slider caused the scale of the face to increase, and increasing the radius slider increased the number of faces that were magnified.

This method increased the resolution so that the individual genes could be seen more clearly. However, texture mapping only allowed a certain number of pixels in the texture map, so each gene could be represented with only a small number of pixels before the maximum resolution was reached. To fix this, I created a separate square mesh and texture map for a zoom window in the position where the face was and kept track of which genes were in each face. I then texture mapped the genes that had been in the face onto the square mesh. The genes were much clearer, and I was able to put each gene in its correct location relative to the other genes and label all of the genes. The zoom window stays at the top of the circular mesh, and when the user uses the clockwise or counter-clockwise buttons, the whole mesh is rotated and a new face is put into zoom window. Figure 2 shows the status of the project.

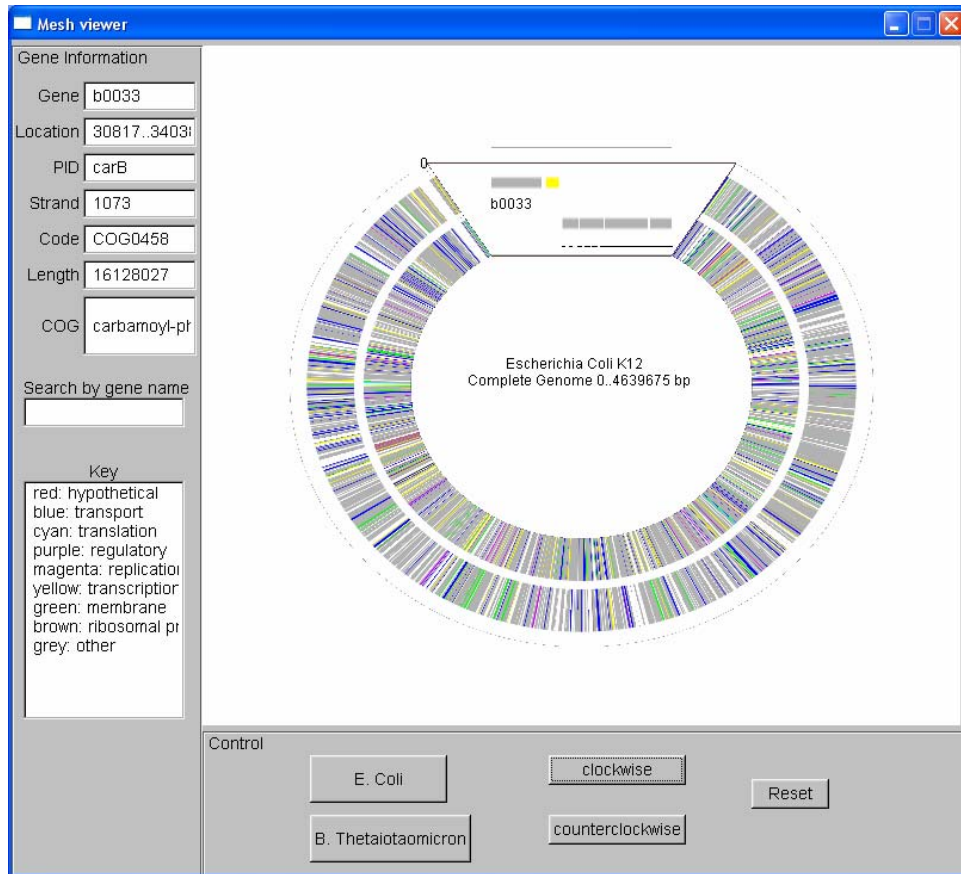


Figure 2: Final Status of project

IV. Results and Conclusions

This tool should be very useful for a biologist trying to locate a specific gene in the E. Coli or B. Theta genomes and visualize its position and length within the genome. The user interface is simple and easy to use. Overall, it was successful because a biologist should find it much easier to analyze the genome by viewing the map than by reading a large file of genome data.

V. Future Work

The next step will be to show the tool to biologists studying this data and see if it will be useful in their research and what other functions are needed. It is also necessary to verify the information so that all of the genes are in the correct position and the information displayed corresponds to the correct gene. It will also be useful to enhance the tool so that it can read data from any genome. Once all functions are implemented, the tool should then be translated into a web-based language for easier access.

VI. References

1. Cavanagh, Sheelagh. "Achieving Higher Manification in Context".
2. "Escherichia coli K-12 MG1655 [genomeMapTool, ver.2.3]".
<http://www.cifn.unam.mx/Computational_Genomics/regulondb/HtmlDocs/ECK12_ZoomTool.html> June 2005.
3. "From genes/molecules to pathway".
<<http://www.genome.jp/kegg/docs/slides/tutorial2/gp04.html>> June 2005.
4. "UCSC Genome Browser".
< <http://genome.ucsc.edu/cgi-bin/hgTracks?clade=vertebrate&org=Human&db=hg17&position=chr7:127,471,196-127,495,720&pix=620&hgsid=58863947&Submit=submit>> July 2005.