

Product Vision

Group 101010

Tycho van Heems - tvheems - 4464567
Dex van Leeuwen - dqvanleeuwen - 475461
Martijn Straatman - martijnstraatm - 4442504
Jochem Tiessen - jtiessen - 4478339
Michael Tran - michaeltran - 4499638

May 18, 2017

Contents

1	Introduction	3
2	Target Customer	4
3	Customer Needs	5
4	Comparison to other products	6
4.1	Cytoscape	6
4.2	Bandage	6
4.3	Our application	6
5	Timeframe and Budget	7
5.1	Timeframe	7
5.2	Budget	7
6	Glossary	8

1 Introduction

To help with research on DNA, there are certain tools which can display graphs represent DNA. Currently, there are not enough tools yet to efficiently work with genomes. Genomes contain a lot of information, which makes it easy to miss important details when studying a genome. Especially when comparing multiple different genomes, it may be difficult to keep track of those, let alone retrieve useful information from these genomes.

Our goal is to provide a new tool for studying DNA. We will make a multi-genome browser, which can read in files containing the information of the genomes. This information is contained within a graph. This graph or parts of it should be visualized on screen. The user of the multi-genome browser will be able to retrieve subgraphs which can be zoomed in on and zoomed out of. It will also be possible to pan the screen to see neighboring parts of the subgraph. Another important aspect of the browser will be the ability to lookup genomes and annotate them.

The rest of this document will explain how we (the development team) envisions the product by talking about our customer and what the customer wants. We will also talk about how our product differs from already existing tools. Lastly, the timeframe and budget in which the product must be made will be discussed.

2 Target Customer

This chapter will introduce you to our target customer and why the product will be made for the customer.

The genome browser will be made for a company called GenomeViz Inc. GenomeViz Inc. is an aspiring company on DNA analysis. The company researches all kinds of DNA, from the DNA of bacteria to the DNA of a tomato. The progress of the research of GenomeViz Inc. is slow, as they cannot seem to find a way to organize and view the large amounts of information contained in the DNA they are researching.

For this reason, GenomeViz Inc. asked us to create a multi-genome browser giving them a clear overview of the genomes they are studying.

Communication with GenomeViz Inc. is done with four different employees working at the company, namely:

T. Abeel is the CEO of GenomeViz Inc.

T. Mokveld is a data scientist at GenomeViz Inc.

J. Linthorst is the CTO of GenomeViz Inc.

L. Krombeen is a software analyst at GenomeViz Inc.

Each of these contacts is specialized in a specific task around the company, so the information and demands we receive for the application are well rounded to all the fields the company specializes in.

3 Customer Needs

In this chapter, it will be explained what the customer expects from the product and what they do not want to see in the product.

GenomeViz Inc. has tasked us with several requirements for the genome browser they have requested. The following list contains the requirements set by GenomeViz Inc.:

In a meeting, T. Abeel¹ stated: "Data scaling is #1 priority", so the data needs to be read as efficiently as possible and the visualization should not take long, so working with the data can be done efficiently.

The application needs to be capable of reading genome graphs from GFA files. The application should then visualize the data from this file.

Drawing the whole graph at once is not necessary, instead a subgraph of a significant size should be drawn. The edge connection the subgraph to the rest of the graph should be visible as if the rest of the graph is off screen.

When viewing a subgraph, the user should be able to zoom in and out and pan the screen. When zooming a bigger or smaller subgraph should be visible to the user. When panning, a part of the subgraph should be moved out of screen and the new part of the subgraph should appear on the opposite side.

The application should also allow the user to easily and intuitively find information about the genomes in this graph.

The user should be able to extract information from the graph.

There should be a possibility of highlighting parts of the graph to keep track of the genome.

The application should be able to run on the following operating systems: Windows, Apple and Linux.

The shape of the nodes in the graph should be a rectangle which width scales with the length of the sequence that node represents.

All these requirements should be met to satisfy the needs of GenomeViz Inc. If one of these requirements is not met, the product may not be suitable enough to use for GenomeViz Inc. which will not only be a letdown for GenomeViz Inc. but also for the development team.

¹T. Abeel is the CEO of GenomeViz Inc.

4 Comparison to other products

A potential issue is that the product might not be interesting because it doesn't provide users with any kind of functionality that other products don't already provide them with. To avoid this issue, we have researched what kind of alternative applications can be found that provide services similar to ones our application provides.

There are not many products that are designed for this purpose, however there are two notable alternatives. The first is Cytoscape², and the second is Bandage³.

4.1 Cytoscape

Cytoscape is used to visualize networks as graphs, and was originally designed to visualize molecules. However, when using Cytoscape to visualize genomes, it becomes very apparent that its focus is not alignment graphs, as its user interface is unnecessarily complex when used for linear graphs. Because our application is specifically designed for alignment graphs the user interface will be much clearer and more intuitive.

4.2 Bandage

Bandage is used specifically for De Novo assembly graphs, which contain cycles causing the graphs to look cluttered, while our application is only made for alignment graphs, which are linear and clear. The way the graphs of bandage handle the large amounts of data is efficient, which is their strongest point. This is also the reason that Bandage is one of the more popular applications in the field of DNA analysis.

4.3 Our application

The reason our application should be used in comparison to other applications is mainly that our application is focused towards linear alignment graphs of DNA. Besides that, our application will provide a 1-based coordinate system, which can be used to easily go to a specific genome within the graph. Lastly, our application will also provide DNA annotations, which can be used to make sense of DNA sequences. Cytoscape can represent large graphs well, but because it is a tool with many uses, its user interface can be confusing. Our application is focused on loading and visualizing even larger graphs and DNA alignment, which is why there will be less functions in the user interface. Bandage is also focused on DNA analysis, but on a different field in DNA analysis, which is why our application will be the preferred choice for the task it is made for, creating linear alignment graphs.

²Cytoscape. <http://www.cytoscape.org/>

³Bandage. <https://rrwick.github.io/Bandage/>

5 Timeframe and Budget

This chapter discusses the timeframe and budget the development team received to make this application.

5.1 Timeframe

GenomeViz Inc. wants to use the multi-genome browser in ten weeks, which means that the development team has a total of ten weeks to finish the application. It is assumed that the development team puts in a total of 140 hours per week in the product. There will be several meetings throughout the timeframe with the customer, so the development team and the customer can talk about the application and the features it should contain. During those meetings, it is expected that the development team brings demo with new features to show the progress.

5.2 Budget

GenomeViz Inc. assigned a budget of 0, -. This means that the development team must find all resources needed to make the product itself and cannot rely on outside sources.

6 Glossary

Genome - A genome is a piece of genetic information that consists of DNA.

Mutation - A mutation in a genome is a difference in a genome when compared to another genome.

Genome browser - A genome browser is an application that can display the information of genomes and navigate efficiently through that information.

Graph - A visual representation of relations of objects.

Data scaling - Data scaling is the ability of an application to work with amounts of data that are normally too large to handle efficiently.

Timeframe - The timeframe of a project is the amount of time the project needs to be finished in.

Budget - The budget is the amount of money that is made available for the project.

References

Nathans, D. (n.d.). BrainyQuote.com. Retrieved May 10, 2017, from BrainyQuote.com Web site: <https://www.brainyquote.com/quotes/quotes/d/danielnath320033.html>

Shannon, P., Markiel, A. & Ozier, O., et al. (2003). "Cytoscape: a software environment for integrated models of biomolecular interaction networks". *Genome Res.* 13 (11): 2498–504.

Wallace, D.C., Singh, G., Lott, M.T., Hodge, J.A. & Schurr, T.G. (1988). Mitochondrial DNA Mutation Associated with Leber's Hereditary Optic Neuropathy. 242(4884), 1427-1430. doi: 10.1126/science.3201231

Wick R.R., Schultz M.B., Zobel J. & Holt K.E. (2015). Bandage: interactive visualization of de novo genome assemblies. *Bioinformatics*, 31(20), 3350-3352

Zimran, A., Gross, E., West, C., Sorge, J., Kubitz, M., & Beutler, E. (1989). Prediction of severity of Gaucher's disease by identification of mutations at DNA level. *The Lancet*, 334(8659), 349-352.