

Product vision

Byzantine Generals

CTGGTGGTGCTCAGCTGCAAGTCAAGCTGCTCTCTGGGCTGTGATCTCCCTGAGACC
CACAGCCTGGATAACAGGAGGACCTTGATGCTCCTGGCACAAATGAGCAGAATCTCT
CCTTCCTCCTGTCTGATGGACAGACATGACTTTGGATTTCCCCAGGAGGAGTTTGA
GGCAACCAGTTCCAGAAGGCTCCAGCCATCTCTGTCCTCCATGAGCTGATCCAGCAG
ATCTTCAACCTCTTTACCACAAAAGATTCATCTGCTGCTTGGGATGAGGACCTCCTA
GACAAATTCTGCACCGAACTCTACCAGCAGCTGAATGACTTGGAAGCCTGTGTGATG
CAGGAGGAGAGGGTGGGAGAACTCCCCTGATGAATGCGGACTCCATCTTGGCTGTG
AAGAAATACTTCCGAAGAATCACTCTCTATCTGACAGAGAAGAAATACAGCCCTTGT
GCCTGGGAGGTTGTCAGAGCAGAAATCATGAGATCCTCTCTTTATCAACAACTTGC
AAGAAAGATTAAGGAGGAAGGAATAA, TGTGATCTCCCTGAGACCCACAGCCTGGA
TAACAGGAGGACCTTGATGCTCCTGGCACAAATGAGCAGAATCTCTCCTTCCTCCTG
TCTGATGGACAGACATGACTTTGGATTTCCCCAGGAGGAGTTTGATGGCAACCAGTT
CCAGAAGGCTCCAGCCATCTCTGTCCTCCATGAGCTGATCCAGCAGATCTTCAACCT

Preface

This document is created for answering the following questions:

1. Who is the target customer?
2. Which customer needs will the product address?
3. Which product attributes are crucial to satisfy the selected needs, and therefore to the success of the product?
4. How does the product compare against existing products, both from competitors and the same company? What are the product's unique selling points?
5. What is the target timeframe and budget to develop and launch the product?

Each of these questions will be answered to their corresponding section.

Ali Smesseim, 4386248, asmesseim
Samuel Sital, 4225139, ssital
Kamran Tadzjibov, 4280784, ktadzjibov
Ravi Autar, 4361172, raviautar
Adam el Khalki, 4348435, aelkhalki
Byzantine Generals
Delft, April 2016

Contents

1	The customer	1
1.1	Background information	1
1.2	Facts about Broad Institute	1
2	The Customers needs	2
2.1	Problem of the customer	2
2.2	Addressed needs	2
3	Product attributes	3
3.1	Crucial attributes	3
4	Uniqueness	4
4.1	Main improvement	4
5	Timeframe	5
6	Glossary	6

The customer

1.1. Background information

The product, addressed in chapter 2, is meant for **Broad Institute of MIT and Harvard**. Broad Institute is the world's leading biomedical research institute dedicated to the mission of using the full power of genomics to transform the understanding and treatment of disease. They try to describe all of the molecular components of life and their connections; discover the molecular basis of major human diseases; develop effective new approaches to diagnostics and therapeutics; and disseminate discoveries, tools, methods, and data openly to the entire scientific community. It was founded in 2004 by Eric Lander and Eli and Edythe L. Broad.

1.2. Facts about Broad Institute

- The Broad Institute has led international projects to create the databases and tools that have made possible systematic studies of the genetic basis of disease, and shared them freely with the scientific community.
- The Broad Institute has a repository of more than 1.3 million biological samples.
- Broad Institute has a couple areas of focus including cancer, genome sequencing and analysis and infectious disease.

2

The Customers needs

To describe which customers needs the product is going to address, we have to evaluate the needs of the customer in general.

2.1. Problem of the customer

DNA is the universal code of life and is represented by the four letters A, T, G and C. The genome is the blueprint and the source code of the operating system of an organism, whether it is a human, a plant or a bacterium. Mutations in the genome will result in changes in how genes work. Small differences between the genomes of people are responsible for many traits we use to differentiate people, such as eye color, but they are equally responsible for differences in our immune systems, allergies, the propensity to get cancer, etc.

Technological advances over the last few years have enabled scientists to determine the DNA sequence of thousands of organisms, including the human genome, but also many bacterial organisms. Bacteria are an essential part of who we are. The human body harbors more bacterial cells than 'real' human cells. While there are bacteria that perform various useful tasks for us, there are also plenty of bacterial species that make us sick. Bacterial infections are responsible for millions of deaths each year, and most people will be treated with antibiotics at some point in their lives to fight off an infection. Just like in human cells, spontaneous mutations also occur in bacterial cells. Some of these mutations will have negative effects, but others may make the bacteria immune to antibiotics. From the bacteria's point of view this is a very desirable feature, but from the medical perspective this is a serious problem as it limits our ability to treat patients.

Even though sequencing technology has enabled scientists to determine the DNA sequence of thousands of bacteria, there is a distinct lack in computational tools to interpret these big data sets of genomes.

2.2. Addressed needs

To address the need of tools to interpret these big data sets of genomes, the product should be an interactive visualization tool that will enable us to explore the genome content of multiple bacteria. In chapter 3 we will describe what attributes the product must have.

3

Product attributes

3.1. Crucial attributes

The product has a couple of crucial attributes that will guarantee the success of the product. The product should

- enable us to explore the genome architecture of multiple strains exploring using a sequence graph.
- put this graph in the context of the evolutionary relationship between bacteria
- put bubbles (mutations) in the graph in the context of well-known reference genomes with their gene annotations and integrate with other reference databases
- provide semantic zooming to enable useful visual interpretation at various zoom levels from whole-genome to individual mutations
- have indications for convergent evolution of variants
- identify mutations and determine the type of variant (insertion, deletion, SNP) uniformly across the samples
- have visual encoding for different classes of mutations and the ability to filter on mutation class
- integrate with other resources, such as literature databases, mutation databases, to identify graph features that are interesting for further investigation.
- provide visual representation and encoding of meta-data associated with samples, such as drug resistance, location of isolation, isolation date, etc.

4

Uniqueness

As there are not a lot of these tools out there, it is almost impossible to compare the product with any similar product. However the product improves on some key aspects of existing tools like our product.

4.1. Main improvement

Most of the current visualizers use a very deeply integrated GUI. This integrated part of the tool causes the tool to be very inflexible. For example the tool is not portable at all. Furthermore sharing gets very complicated as the model-view-controller-model gets violated if you do so.

The product enforces the model-view-controller-model by using the client-server-model. Instead of integrating the GUI into the tool, the product relies on a web-application to visualize the data generated by the server. This ensures the tool is very portable as all you need to access the tool is a browser.

5

Timeframe

The timeframe the product must be developed in is 9 weeks. This time is divided in 9 sprints of 1 week long. Each sprint has a certain amount of work that must be done in that specific sprint.

6

Glossary

D

- **DNA (deoxyribonucleic acid)** is a self-replicating material which is present in nearly all living organisms as the main constituent of chromosomes. It is the carrier of genetic information.

M

- **Mutation** is the changing of the structure of a gene, resulting in a variant form which may be transmitted to subsequent generations, caused by the alteration of single base units in DNA, or the deletion, insertion, or rearrangement of larger sections of genes or chromosomes.

P

- **Phylogenetic tree** is a branching diagram or "tree" showing the inferred evolutionary relationships among various biological species or other entities—their phylogeny—based upon similarities and differences in their physical or genetic characteristics.

S

- **Semantic zooming** is a form of details on demand that lets the user see different amounts of detail in a view by zooming in and out.