[DRAFT] Product vision

Group PL4
28 April 2016

Arthur Breurkes	4398033
Ricardo Jongerius	4379500
Niek van der Laan	4296915
Daphne van Tetering	4375165
Niels Warnars	4372069
Ties Westerborg	4386817

Context coordinator: Dr. T. Abeel
Teaching assistants: Mr S. Huisman MSc
Mr J. Linthorst MSc
Mr J. Hommes

Table of Contents

1. Introduction

The discovery and manufacturing of dozens of new antibiotic medicines back in the 20th century has started a revolution in the treatment of patients and has drastically reduced the mortality rate for several bacterial infections that have now become treatable. This revolution did not come without much effort. Years of research have contributed to the development of these new antibiotics. But drug resistance is on the lure as bacterial gen mutations threaten to cause a rise in the number of treatment failures. [1] If drug resistant bacteria are spread humanity might lose its ability to treat conditions that are not considered easily treatable.

In an attempt to develop alternative treatment methods for infections caused by resistant bacteria scientists are looking more and more at research fields as microbial genomics for help in understanding the mutations that cause drug resistance. [2] This quest asks for large sets of samples genome data coming from multiple infected patients which causes a bottleneck. As the amount of research data grows so does the amount of time to analyse this data, therefore suitable visualisation and analysis software is needed that can process several hundreds or thousands of genome samples at a time.

2016’s Programming Life Context Project focusses on addressing this need by introducing tooling that can be used to interactively explore a sequence graphs of different sized sections of bacteria's genome. By putting the graph in the context of the evolutionary relationship of different samples mutations can be traced back till the point in history where it arose.

2. Target audience

The the main audience of the DART-N genome visualizer are researchers working in the field of bioinformatics that specialize in genome research. Code of the project will be made available for the general public. The initial commissioning has been done on behalf of:

- The Broad Institute of MIT and Harvard, a US-based biomedical research center specialized in mapping the human genome, analyzing the molecular basis of infection diseases and mutations responsible for different cancer types. [3][4]
- The KwaZulu Natal Research Institute for Tuberculosis & HIV (KRITH), a South-Africa-based science research institute focussed on translate scientific findings involving Tuberculosis and HIV into new controls to stop the TB and HIV pandemics.[5][6]

3. Customer needs

The main research question that the customers want to have answered is: *How do we speed up the process of scientific discovery in the context of comparative genomics?*

From this research question it can be derived that the customers want an application which can visualize DNA samples of multiple organisms from the same species in a single graph architecture. A large scale visualization might help in performing exploratory trend analysis on a given data set in future research.

This wanted visualization introduces several challenges that are solved when the product is delivered. These challenges all involve visualizing interesting mutations in specific genes for example responsible for drug resistance. The main challenges include:

- *How do genomes with drug resistance genes compare to genomes without these mutations?*
- *Which mutations are present in the genome set?*
- *Do identified mutations result in drug resistant phenotypes?*
- *Can a common ancestor be identified for these mutations?*

4. Project attributes

The most import basic attribute of the product is the ability to interactively explore the genome graph of multiple DNA samples. In order to properly navigate this genomegraph, well functioning semantic zooming that correctly visualizes the graph at various zoom levels is essential. Furthermore, the graph should visualize the evolutionary relationship between different samples in the genomegraph. To prevent clutter and to keep an overview of the compared genomes different mutation classes also have a visual encoding and can be completely filtered away. If mutations are not filtered away they are displayed as bubbles on higher zoomlevels.

5. Product comparison

Two groups of products exist that are competing with our product:

- *TU Delft student developed visualizers*
- *Freely available third party visualizers*

Compared to other TU Delft student developed visualizers all delivered products of previous year's programming life context project [7][8][9][10][11] have a strong set of features and all have several unique advantages over the other products. In our opinion they still all have features that have been implemented a wrong way.

What all of last year's products miss is proper semantic zooming. Basic zooming with bubble collapsing is implemented in all products in a some form, but or the entire genome itself is already poorly visualized or a visualization has to jump to a new level after a zoom action in which suddenly many more nodes were shown.

The number of products that is capable of visualizing genome graphs is limited. The most promising one that has been identified is Bandage (a Bioinformatics Application for Navigating De novo Assembly Graphs Easily).[12] Bandage is able to load a GFA file and visualize it as a graph, although it cannot distinguish between separate genomes in the graphs and it also cannot deal with large genome sets.

Unique selling points

The most unique selling point of the DART-N genome visualizer is its optimized semantic zooming. Whereas other products only contain partial zoom functionality. DART-N will be able to show a smooth transition during zoom operations. In the fully zoomed in state the individual nodes and edges are clearly visible, when a user starts to zoom out nodes and edges that will be collapsed become smaller till they are no longer visible. The nodes that will not collapse will stay the same size

in this case. At the higher zoomlevels individual nodes are no longer the main focus and only their not collapsed edges will be shown to form smooth ribbons.

6. Resource allocation

The timeframe allocated for the development of the product is 10 weeks. In these ten weeks from 18 April till 24 June 1700 development hours have been reserved to create multiple versions using rapid application development allowing prototypes to be delivered in a quick manner. [13]

In the ten week's period every week is concluded with a working prototype.

7. Bibliography

[1] *History of Antibiotics - heeve.com* [Web page] publication date unknown [Retrieved 28-04-2016] Available from: http://www.heeve.com/modern-history/history-of-antibiotics.html

[2] *Whole-genome sequencing targets drug-resistant bacterial infections - Punina et al. 2015* 5-08-2015 [Retrieved 28-04-2016] Available from: http://humgenomics.biomedcentral.com/articles/10.1186/s40246-015-0037-z

[3] *Who is Broad? - The Broad Institute* [Web page] publication date unknown [Retrieved 24-04-2016] Available from: https://www.broadinstitute.org/what-broad/who-broad/who-broad

[4] *Areas of Focus - The Broad Institute* [Web page] publication date unknown [Retrieved 24-04-2016] Available from: https://www.broadinstitute.org/what-broad/areas-focus/areas-focus

[5] *What is K-RITH? - K-RITH* [Web page] publication date unknown [Retrieved 24-04-2016] Available from: http://www.k-rith.org/what-is-k-rith

[6] *Why TB and HIV? - K-RITH* [Web page] publication date unknown [Retrieved 24-04-2016] Available from: http://www.k-rith.org/what-is-k-rith/why-tb-and-hiv

[7] *dnapp - Context 2014-2015 PL1* [Web page] 26-06-2015 [Retrieved 27-04-2016] Available from: https://github.com/AbeelLab/dnapp

[8] *geex - Context 2014-2015 PL2* [Web page] 26-06-2015 [Retrieved 27-04-2016] Available from: https://github.com/AbeelLab/geex

[9] *Helix^2 - Context 2014-2015 PL3* [Web page] 25-06-2015 [Retrieved 27-04-2016] Available from: https://github.com/AbeelLab/helix2

[10] *DNAinator - Context 2014-2015 PL4* [Web page] 25-06-2015 [Retrieved 27-04-2016] Available from: https://github.com/AbeelLab/dnainator

[11] *Life Tiles - Context 2014-2015 PL5* [Web page] 30-06-2015 [Retrieved 27-04-2016] Available from: https://github.com/AbeelLab/LifeTiles

[12] *Bandage - Ryan Wick* [Web page] 14-02-2015 [Retrieved 28-04-2016] Available from: https://rrwick.github.io/Bandage/

[13] *What is RAD model- advantages, disadvantages and when to use it? - istqbexamcertification.com* [Web page] publication date unknown [Retrieved 24-04-2016] Available

from:
http://istqbexamcertification.com/what-is-rad-model-advantages-disadvantages-and-when-to-use-it/