Product Planning

Group PL-4

TI2806 Contextproject
Programming Life
Group PL-4

TU Delft

Product Planning

Programming Life

by

Group PL-4

Arthur Breurkes
Ricardo Jongerius
Niek van der Laan
Daphne van Tetering
Niels Warnars
Ties Westerborg

Project duration: April 18, 2016 – July 1, 2016

TU Delft

Contents

1 Introduction

The discovery and manufacturing of dozens of new antibiotic medicines back in the 20^{th} 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][Heeve, n.d.] If drug resistant bacteria are spread, humanity might lose its ability to treat conditions that are not considered easily treatable.

The client, The Broad Institute, is a biomedical and genomic research center. 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][Punina, 2015], as does Broad Institute. This quest asks for large sets of samples of 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 focuses on addressing this need by introducing tooling that can be used to interactively explore sequence graphs of different sized sections of genomes of bacteria. 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 result of this project is a genome visualiser that can be used to interactively compare genomes of DNA samples coming from multiple organisms. A researcher working on mutations in the genome of multiple organisms can use the product to view the exact type of mutation that has occurred in the DNA of a sample compared to the reference. The following chapter describes the product backlog and the roadmap of our visualiser. The backlog provides an overview of basic milestones that have to be reached after each weekly delivery. The roadmap in turn provides an overview of which functionalities are expected to be included in each release version.

2.1. High-level product backlog

In this section, Epic user stories are made use of to describe the list of backlog items. An epic has the same format as a normal User story, it only includes a bigger scope.[3][Scrum, n.d.] Our epics are as follows:

- As a user, I want to be able to navigate through the sequence graph in an intuitive and free manner.
- As a user, I want to be able to zoom in onto the graph, allowing me to see more details as I am zooming in.
- As a user, I want to be able to see the corresponding phylogenetic tree and be able to modify my graph through selections of the leafs.
- As a user, I want to be able to view the percentage of mutations, deletions or insertions in a pie chart and to view the percentages of A, T, C, G basic in a pie chart.
- As a user, I want to be able to see a representation of meta-data that is associated with the samples, such as drug-resistance, the location of the isolation and the isolation date.

2.2. Roadmap

- `Week 1`
  Design decisions are made, the first graphical interface is designed and the first data set is analysed. At the end of the week, the first demo is finished. This means that the DNA researcher will be able to load a small genome and analyse it.
- `Week 2`
  The DNA researcher will now be able to use a first concept of semantic zooming that collapses nodes as they are zoomed in on.
- `Week 3`
  The DNA researcher will be able to and view the phylogenetic tree of the used genome and scroll through the phylogenetic tree.

- `Week 4`
  The DNA researcher will be able to load more genomes (100+) at once, and compare these by picking a reference genome.
- `Week 5`
  The DNA researcher will be able to fully use semantic zooming. This includes cluster collapsing, tracing paths of individual genomes and information about the type of mutation.
- `Week 6`
  The DNA researcher will be able to view the percentage of mutations, deletions or insertions in a pie chart and will be able to view the percentages of A, T, C, G bases in a pie chart.
- `Week 7`
  The DNA researcher will be able to see a representation of meta-data that is associated with the samples, such as drug resistance, the location of the isolation, the isolation date, etc.
- `Week 8`
  The DNA researcher will be able to see an indication for the convergent evolution of several variant.

3

User Stories

In the provided user stories a high-level definition of features or characteristics of our visualiser is described. These user stories include both functional and non-functional specifications of our visualizer.

- `Data loading`
  As a user
  I want to be able to manually load a Graphical Fragment Assembly (GFA) data file
  So that any given set of genomes can be loaded and transformed into a graph
- `Large number of genomes`
  Given that a GFA file of 340 genomes is given
  Then it should be loadable and processable on a low-end desktop machine running a dualcore processor with 4GBs of RAM within three seconds.
- `Navigation`
  As a user
  I want to be able to explore the comparison graph
  So that mutations and evolution between samples can be studied
- `Nucleotide-level comparison`
  A a user
  When zoomed in close enough
  that individual nucleotides can be distinguished I want to be able to see the affected sequence of nucleotides.
- `Whole-genome comparison`
  As a user
  When the graph visualisation is fully zoomed out
  I should be shown ribbons with indistinguishable vertices instead of a detailed graph
- `Evolutionary relationship between bacteria`
  As a user
  When zoomed in on a section where mutations exist
  I should be able to see the corresponding evolutionary information in the graph.
- `Mutation visualisation`
  As a user
  I want to be able distinguish different types of mutation

  - `Scenario 1:  Bubble visualisation`
    Given that a genome contains a mutation and the graph is zoomed in to the level at which individual nucleotides can be distinguished
    Then mutations should be shown as bubbles which distinguish insertion, deletion, duplication and inversion.

- `Scenario 2: Glyphs`
  Given that a genome contains a common mutation
  Then this mutation should be visualized using a glyph if the graph is zoomed out too far to distinguish individual nucleotides.
- `Scenario 3: Mutation class filtering`
  Given that multiple classes of mutations exist within a set of genomes and I am only interested in a certain class of mutations
  Then it should be possible to filter on this specific mutation class

4

Definition of Done

The definition of done is created so that everyone that is involved in the project knows what done means. This also means that if a feature is done, everyone is sure that there is nothing left to do and the feature can be closed. We divided our DoD in three parts, all concerning a different phase of the project.

4.1. Backlog Items

We consider a backlog item to be done when to following requirements are met:

- All user stories are met
- Functional tests are performed by all members of the team
- Unit tests are written and passed
- The code is commented
- The build is passing

4.2. Sprints

We consider a sprint to be done when to following requirements are met:

- All user stories and features indicated in the sprint are included
- Unit tests are written and passed
- The build is passing
- The code is commented and the documentation is written and updated

4.3. Release

We consider a Release to be done when to following requirements are met:

- The demo is handed over to the customer
- The documentation is included with the demo
- Product has passed the User Acceptance Test

5

Glossary

- DoD - Definition of Done
- Epic - Epic User Story

Bibliography

[1] Heeve. History of antibiotics. *http://www.heeve.com/modern-history/history-of-antibiotics.html*, n.d.

[2] Punina. Whole-genome sequencing targets drug-resistant bacterial infections. *http://humgenomics.biomedcentral.com/articles/10.1186/s40246-015-0037-z*, 2015.

[3] Scrum. Scrum begrippen. *http://www.scrum.nl/site/Scrum-Begrippen-agile-scrumEpic*, n.d.