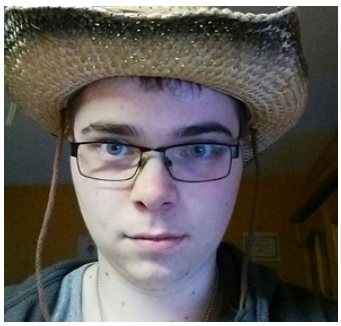
Product Vision

Context project - Programming life

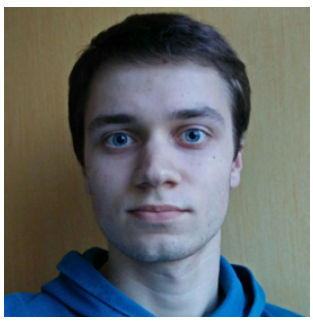
Justin

Marissa

Mark



Chak Shun
Scrum Master



Maarten
Product Owner

Team PL1

Chak Shun Yu	csyu	4302567
Justin van der Hout	jrtevanderhout	4319982
Mark Pasterkamp	mpasterkamp	4281500
Marissa van der Wel	mmvanderwel	4323602
Maarten Flikkema	mhflikkema	4306538

Abstract

This document represents the developers' vision on the project context. We will define our vision by answering the fundamental questions according to Roman Pichler (Pichler, R., 2009). We will explain who we think our customers are, what they would need from our product, and which product attributes are essential to satisfy those needs. Furthermore, we will compare our product vision with competing software. Finally, we will talk about the target timeframe and resources we had available to develop and launch our product.

Table of contents

Introduction

The goal of this Context Project is to develop a program which provides an interactive visualization of DNA sequence graphs. This visualization will represent the genome architecture of organisms of interest for the user, such as drug-resistant human pathogens. The users of this program should be able to interact with the visualization of a large quantity of genomes, and extract useful data. With this they can formulate novel hypotheses, check existing ones, and identify outliers, trends and patterns in the data. (Abeel, T., 2015)

The customer

The two main parties involved with our project as customers are the Broad Institute of MIT and Harvard, and KwaZulu Natal Research Institute for Tuberculosis & HIV (KRITH) (Abeel, T., 2015).

The Genomics Platform of the Broad Institute of MIT and Harvard is a genome sequencing center located in Cambridge, Massachusetts, United States of America interested in exploring large genomic datasets, specialized in sample handling, genotyping, gene expression, and genome sequencing (“Genomic Platform”, 2015).

KRITH is a TB and HIV research center located in Durban, South Africa interesting in stopping the TB and HIV pandemics in South Africa. “Its mission is to conduct outstanding basic science research on tuberculosis (TB) and HIV and translate the scientific findings into new tools to control TB and HIV.” (“What is K-RITH?”, 2015)

Both parties hope that our program will be a major contribution to moving a step forward in their respective researches.

Customer needs

Some major challenges our program has to tackle are: How do these genomes look in comparison to each other? Which mutations have occurred and where? Have any of these mutations occurred in ‘interesting’ genes? Do these genes associate with phenotypes our customers care about, such as drug resistance? Do mutations share a common ancestor? (Abeel, T., 2015)

Specifically, the program has to satisfy the following needs of the customer (Abeel, T., 2015):

- Enable the user to interactively explore a sequence graph representing the genome architecture of multiple strains.
- Provide semantic zooming to enable useful visual interpretation at various zoom levels from whole-genome to individual mutations. Zooming will be possible in two dimension: Both on sequence axis as on sample axis.
- Put the sequence graph in the context of the evolutionary relationship between bacteria. Using phylogeny we can provide information on how strains are related, which is an important context.
- Identify mutations and determine the type of variant (INDEL, SNP) uniformly across the samples.

- Put mutations in the graph in the context of well-known references genomes with their gene annotations and integrate with other reference databases. Since Mycobacterium tuberculosis is already well studied, we can try to leverage that information and provide it to our users.
- Provide indications for convergent evolution of variants.
- Have visual encodings for different classes of mutations and the ability to filter on mutation class (Insertion, Deletion, SNP, Duplication, Inversion, Translocation).
- Integrate our program with other resources, such as literature databases, mutation databases, to identify graph features that are interesting for further investigation.
- Provide visual representations and encoding of metadata associated with samples, such as drug resistance, location of isolation, isolation date, etc.

Beside these functional requirements, there are three non-functional requirements to be mentioned:

- Language: Java
- Platform: OS independent desktop application
- Source needs to be kept in version control

Crucial product attributes

To satisfy the needs of our customer we will list attributes that will be crucial in satisfying their needs and ascertain success of our product.

Our product will provide different semantic zoom levels for users in two different dimensions. Firstly, the user can filter the visual representation on certain genomes or certain parts of the genome sequence graph as the user wishes. Secondly, our product will provide multiple kinds of visual representation in a more biological context. These visual representations will be based on the different stages of DNA: nucleotides, amino-acids and genes. Together, these semantic zoom levels will provide the user with more perspectives and ways to analyze and compare the genomes.

Another crucial attribute of our product is that it will be an OS independent desktop application. To use our product, no (research) data has to be uploaded or put online. All the data regarding the genomes can be kept locally and private for personal use.

The last crucial attribute of our product is that it will try to interact with the user instead of just providing the user with the DNA sequence information. Our product will try to interpret the served data and help the user with analyzing the information. This is done by identifying mutations and determining the type of variant (Insertion, Deletion, SNP, etc.), highlighting mutations that occurred in interesting parts of the genome and providing more relevant data from other sources on the location of the mutation. We want to make our product more than just another genome browsing tool and actually assist the user in their analysis and research.

Comparison with existing products

We have found several genome browsers available on the internet, the one most commonly used being Ensembl. The biggest disadvantage of Ensembl as well as other genome browsers such as GBrowse (Stein, L., 2011) and xGDB (“xGDB: The eXtensible Genome Data Broker”, 2006) is that they are web based applications. This results in a slow user experience, the need to have an internet connection, and finally the need to upload the genomes, which imposes the risk of confidential information being compromised. Furthermore, the data uploaded to Ensembl will be made freely available online. (“Ensembl Genome Browser”, 2015)

In 2007, MaizeGDB, an informatics service, distributed a survey about the use of genome browsers. “The survey results indicate that cooperators do not consider Ensembl easy to use, and it is definitely perceived to be slow when compared to the other software available.” (Sen, T.Z., et al., 2010)

In the field of genome browsers, the three most desired features are usability, visuals, and speed (Sen, T.Z., et al., 2010). One of the advantages of our product compared to competitors is that it is an offline application, meaning that no internet connection will be required to use any of its features. This will guarantee that our application is also available offline and more responsive since there is no network lag.

Target timeframe & Resources

The final product has to be fully developed and launched within 10 weeks starting at the kickoff on April 20, 2015. There will be 8 iterations following the scrum methodology.

During this time, we will work as a team consisting of five developers to make this product a success.

Bibliography

- Abeel, T. (2015). *TI2806 - Contextproject Programming Life* [slides].
  Retrieved from: https://blackboard.tudelft.nl TI2806/Contexts/Programming Life/
- Abeel, T. (2015). *Context-project: Programming Life* [requirement document].
  Retrieved from: https://blackboard.tudelft.nl TI2806/Contexts/Programming Life/
- “Ensembl Genome Browser” [homepage] (2015)
  Retrieved from: http://www.ensembl.org (8-5-2015)
- Genomics Platform (2015)
  Retrieved from:
  https://www.broadinstitute.org/scientific-community/science/platforms/genomics/genomics-platform
- Pichler, R. (2009). *The Product Vision* [article]
  Retrieved from: https://www.scrumalliance.org/community/articles/2009/january/the-product-vision
- Sen, T.Z., Harper, L.C., Schaeffer, M.L., Andorf, C.M., Seigfried, T.E., Campbell, D.A., Lawrence1, C.J. (2010). *Choosing a genome browser for a Model Organism Database: surveying the Maize community*.
  Retrieved from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2911842
- Stein, L. (2011). *GBrowse/README at master*
  Retrieved from: https://github.com/GMOD/GBrowse/blob/master/README
- “What is K-RITH?” (2015)
  Retrieved from: http://www.k-rith.org/what-is-k-rith
- “xGDB: The eXtensible Genome Data Broker” (2006)
  Retrieved from: http://xgdb.sourceforge.net