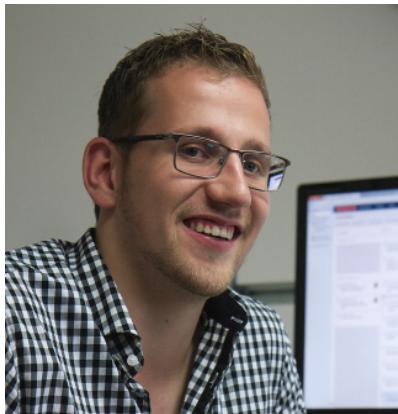


LifeTiles - Product Vision



Jos Winter



Joren Hammudoglu



Rutger van den Berg



Albert Smit



Arjan Langerak

Inhoudsopgave

1. Product

- 1.1. Product attributes
- 1.2. Product comparison and USP's
- 1.3. Target Timeframe and Budget

2. Customer

- 2.1. Customer needs
- 2.2. Product attributes and customer needs correlation

1. Product

LifeTiles is an application which makes it possible for its user to open a set of phylogenetic tree files and DNA sequence graph files. LifeTiles is a tool for the interactive visualization of DNA sequence graphs to represent the genome architecture of organisms of interest. We made a mockup, Figure 1, to illustrate our ideas for the Data Viewer, Tree Viewer and Graph Viewer and to use as a reference while implementing our application. With the parsed data the program will be able to visualize several datasets:

- A **Data Viewer** which shows all sequence/sample names so the user has an overview of the data which is currently viewed by the program.
- A **Tree Viewer** which shows the contents of the Newick files in the structure of a Sunburst diagram. This shows the relation between the different samples used in the program.
- A **Graph Viewer** which shows the contents of the Node and Edge files in the structure of a Tile/Grid diagram. It visualizes multiple DNA strains and the mutations between those DNA strains. Similar mutations and DNA parts are grouped in a larger block. Insertions in the non-reference sequence will create an empty block inside of the reference sequence so the DNA strains will stay aligned. This way we can achieve a reasonable overview of small and large datasets of DNA strains. This process is clearly observable in Figure 1.

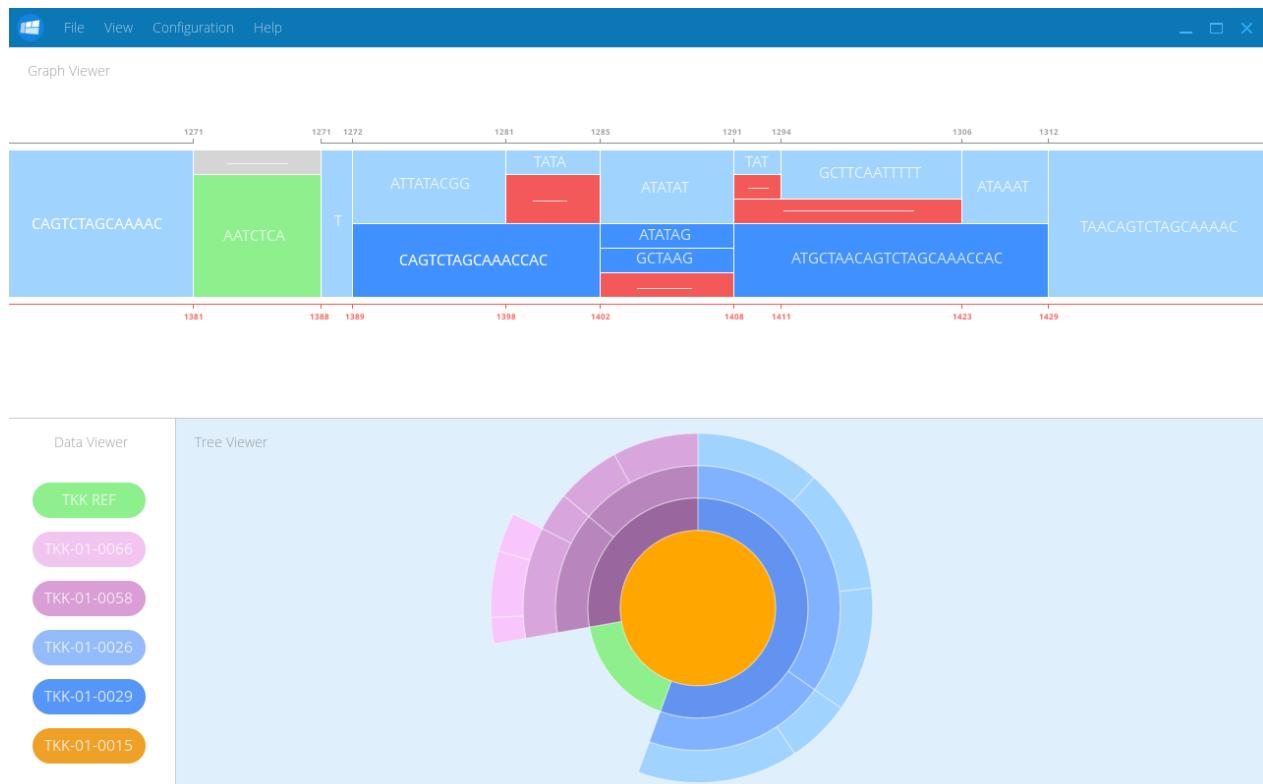


Figure 1: Mock-up of the LifeTiles application.

1.1. Product attributes

Attributes of our product are:

1. We provide an innovative and intuitive way of displaying the graph data using a tile/grid diagram.
2. We make semantic zooming possible by filtering out the least important data and indicating the loss of data, interesting details can be highlighted using indications.
3. We provide a clear and well structured overview of the phylogenetic tree using a Sunburst diagram.
4. We provide the ability to filter on subgroups of the phylogenetic tree which will apply the filter to the graph viewer.
5. The standard way in which we will display mutations will use mutation highlighting which displays each type of mutation in a certain color. (insertion: green, deletion: red etc.)
6. We provide a way to easily filter the mutation highlighting used in the graph viewer. (mutation type, minimal mutation amount)
7. We provide a display of variant indications in the phylogenetic tree in the Sunburst diagram by naming the nodes in the Sunburst diagram to the correct variant.
8. We provide a Data Viewer tableview which will contain the sequences/samples and metadata used by the user.
9. We provide a good overview of the data by using a clean interface design.
10. We provide the use of external resources by collecting external meta-data about the used samples and by applying it to the data viewer and the tree viewer.

1.2. Product comparison and USPs

There are several existing products which provide a visualization of DNA sequences, but there aren't a lot of products which provide interactive visualization of DNA sequence graphs. The following set of products provide visualization of DNA sequences:

- CLC Sequence Viewer provides a large number of bioinformatics analysis tools. It provides visualization, creation and editing of alignments and the visualization of phylogenetic trees. Its unique selling point is that it's specialized in visualization of DNA sequences, phylogenetic trees and editing.
- Artemis: Genome Browser and Annotation Tool is a free genome browser and annotation tool that allows visualisation of sequence features, next generation data and the results of analyses within the context of the sequence, and also its six-frame translation. Its unique selling point is that it allows specialized visualization of DNA sequences, annotations and sequence features.

The following product provides visualization of DNA sequence graphs:

- Cytoscape is an open source software platform for visualizing complex networks and integrating these with any type of attribute data. Its unique selling point is that it can work with a lot of different file types. So it isn't specifically designed for the visualization of DNA sequence graphs which decreases its performance.

1.3. Target Timeframe and Budget

The target timeframe for the project is from 20 april 2015 to 19 june 2015 which adds up to 8 weeks of development time in group of 5 persons. We have no budget to develop and launch the product.

2. Customer

The customers can be divided in two groups. The first group of customers are researchers at the Broad Institute of MIT and Harvard in Cambridge, MA, USA. A genome sequencing center interested in sequencing large genomics datasets.

The second group of customers are researchers and doctors at the KwaZulu Natal Research Institute for Tuberculosis and HIV (KRITH) in Durban, South-Africa. A tuberculosis and HIV research center interested in stopping the tuberculosis and HIV pandemics in South-Africa.

2.1. Customer needs

The customer wants a tool for the interactive visualization of DNA sequence graphs to represent the genome architecture of organisms of interest, such as drug-resistant human pathogens. The customer wants to use it for the interactive visualization of large scale pan-genome graphs which enables exploratory data analysis to formulate novel hypotheses, check existing ones, and to identify outliers, trends and patterns in the data. The customer's needs can be described more detailed using the following nine needs:

1. Interactively explore a sequence graph representing the genome architecture of multiple strains.
2. Provide semantic zooming to enable useful visual interpretation at various zoom levels from whole-genome to individual mutations.
3. Put this graph in the context of the evolutionary relationship between bacteria.
4. Identify mutations and determine the type of variant (insertion, deletion, SNP) uniformly across the samples.
5. Put bubbles (mutations) in the graph in the context of well-known references genomes with their gene annotations and integrate with other reference databases.
6. Have indications for convergent evolution of variants.
7. Have visual encodings for different classes of mutations and the ability to filter on mutation class.
8. Integrate with other resources, such as literature databases, mutation databases, to identify graph features that are interesting for further investigation.
9. Provide visual representation and encoding of meta-data associated with samples, such as drug resistance, location of isolation, isolation date, etc.

2.2. Product attributes and customer needs correlation

The customer needs correlate with the product attributes. For every customer need there is at least one product attribute defined which addresses it.

1. To interactively explore a sequence graph representing the genome architecture of multiple strains we provide an innovative and intuitive way of displaying the graph data using a tile/grid diagram.
2. We provide semantic zooming to enable useful visual interpretation at various zoom levels from whole-genome to individual mutations by making semantic zooming possible by filtering out the most important data and indicating the loss of data, interesting details can be highlighted using indications.
3. We put this graph in the context of the evolutionary relationship between bacteria by providing a clear and well structured overview of the phylogenetic tree using a Sunburst diagram which will provide the ability to filter on subgroups of the phylogenetic tree which will apply the filter to the graph viewer.
4. We identify mutations and determine the type of variant (insertion, deletion, SNP) uniformly across the samples by using mutation highlighting which displays each type of mutation in a certain color. (insertion: green, deletion: red etc.)
5. We put bubbles (mutations) in the graph in the context of well-known references genomes with their gene annotations and integrate with other reference databases by ...
6. We have indications for convergent evolution of variants by providing a display of variant indications in the phylogenetic tree in the Sunburst diagram by naming the nodes in the Sunburst diagram to the correct variant.
7. We have visual encodings for different classes of mutations and the ability to filter on mutation class by providing a way to easily filter the mutation highlighting used in the graph viewer. (mutation type, minimal mutation amount)
8. We integrate with other resources, such as literature databases, mutation databases, to identify graph features that are interesting for further investigation by providing the use of external resources by collecting external meta-data about the used samples and by applying it to the data viewer and the tree viewer.
9. We provide visual representation and encoding of meta-data associated with samples, such as drug resistance, location of isolation, isolation date, etc. by providing a Data Viewer tableview which will contain the sequences/samples and metadata used by the user.