

SBOLExplorer

Russell Kennington - u0946645 - arussellk@gmail.com

Jiahui Chen, u0980890 - jiahui.chen@utah.edu

<https://github.com/arussellk/dataviscourse-pr-sbolexplorer>

Background and Motivation

This project is motivated by the research work of Michael Zhang, a mutual friend and student at The U. The data we are visualizing is a hierarchy of synthetic genetic data. The Synthetic Biology Open Standard (SBOL - <http://sbolstandard.org/>) defines a data format which describes genetic building blocks and their relationships. SBOLExplorer will provide a way to view and navigate SBOL data to give researchers deeper understanding and more efficient discovery of their data.

Project Objectives. We want to answer:

- What subparts compose this genetic part?
 - An SBOLExplorer user can search for a part like “BBa_E0041”, see relevant search results, click on a result, and view “BBa_E0041” and its child components.
 - The ability to quickly search and navigate SBOL data will help researchers make better use of their SBOL data.
- How do these subparts work together to make this genetic part?
 - SBOLExplorer’s hierarchical tree will have a genetic sequence visualization which shows where child parts fit together to make a genetic component.
 - Visualizing the size and location of child components in the parent genetic part could help researchers see new relationships.
- What is the detailed information for this genetic part?
 - SBOLExplorer provides both a tooltip and an expanded view for genetic part details.
 - Providing detailed data about a genetic part will let researchers make full use of their SBOL data.

Data.

Our data will be provided by Michael Zhang. Similar data can be found through SynBioHub (e.g. https://synbiohub.org/search/BBa_E0041). If necessary, we will export static JSON files instead of working against a live API.

Data Processing.

We do not expect to perform substantial data cleanup. We may need to manually label the start and stop indices of child components in relation to parent sequences.

For example:

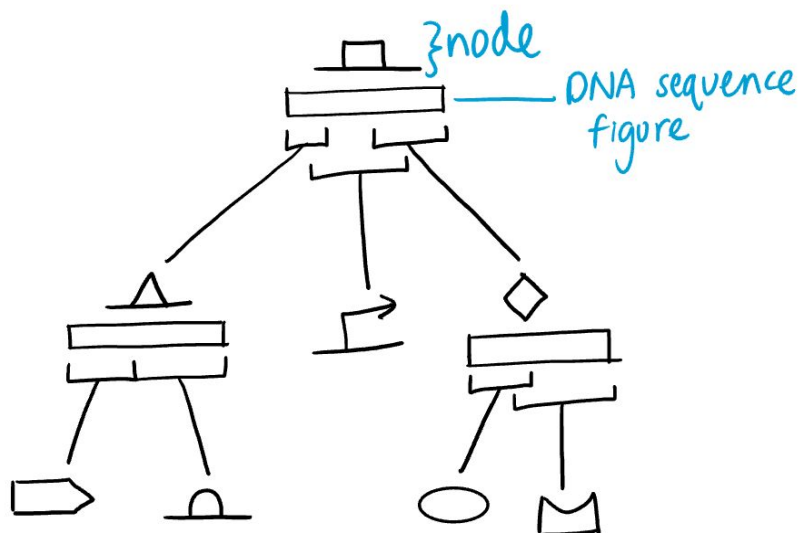
```
{
  sequence: "AAAABBBBCCCC"
  children: [
    {
      sequence: "AAAA",
      parentStart: 0,
      parentStop: 4,
    },
    ...
  ]
}
```

Visualization Design:

We will have a main visualization consisting of a tree with each node representing a genetic part. Each node will be encoded with a quantitative channel that indicates its popularity (page rank). Each edge will represent a dependency: the parent node's DNA sequence includes each child node's DNA sequence. Each node will also show a figure that represents its DNA sequence, and this figure is marked according to which portions of it consist of its children's DNA sequences. The tree will display data dictated by a search dialogue box that searches for genetic parts. Here's a sketch of a broad overview:

Overview:

search	
Name1	<input type="checkbox"/>
Name2	<input type="checkbox"/>
Name3	<input type="checkbox"/>
Name4	<input type="checkbox"/>
Name5	<input type="checkbox"/>

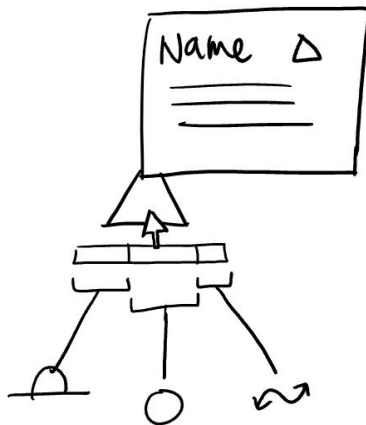


Additional views include:

Popularity indicator (page rank) within the search box, one bar for each genetic part in the search results. This lets users quickly discern if SBOLExplorer has found relevant results and how many of the results are worth viewing.

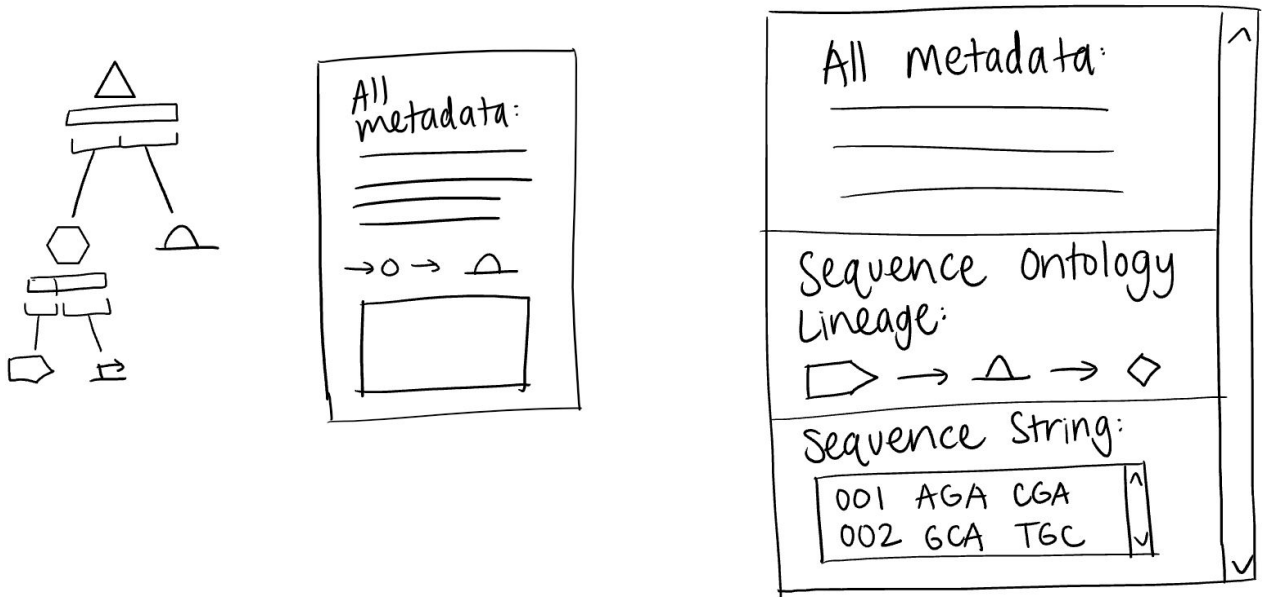
search	
Name1	<input type="checkbox"/>
Name2	<input type="checkbox"/>
Name3	<input type="checkbox"/>
Name4	<input type="checkbox"/>
Name5	<input type="checkbox"/>

Basic info box that displays when hovering over any node/genetic part. The contents of this tooltip may change as we get feedback from future SBOLExplorer users.



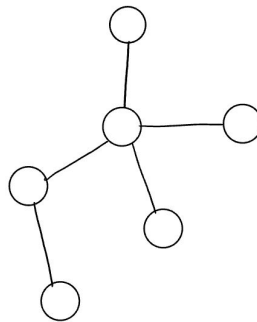
Name	
Description:	
<hr/>	
<hr/>	
<hr/>	

Thorough info box showing all of a genetic part's metadata when the genetic part's node is clicked.



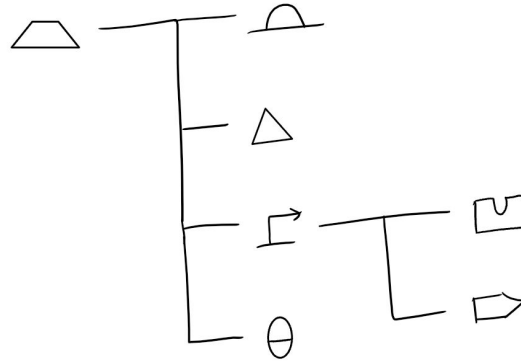
Prototypes:

Graph - The initial idea was to display the data in a simple graph where each node represented a genetic part and each edge represented a DNA sequence dependency:

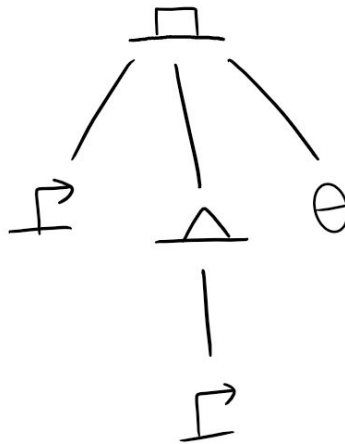


We decided this visualization doesn't convey the most information for this data, as the dependencies should be shown in a more hierarchical manner and should have more influence on the positioning of the nodes.

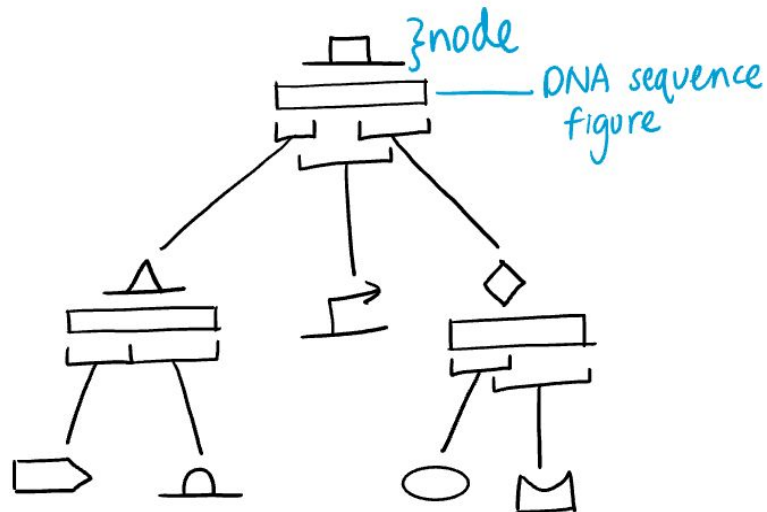
Left to Right Tree - Inspired by filesystem hierarchies, we moved to a left-to-right tree visualization to show dependencies more clearly. We also decided to use the glyphs for each node (the glyphs are part of a field-wild system of symbols for genetic parts and given in the data) instead of one shape for all nodes to show more information on first glance.



Top Down Tree - We decided to do a top-down tree instead of a left-to-right tree because according to Michael a top-down tree would be more readily understood by biologists.



Final visualization design: Top down tree with DNA sequence figure - this is the same as the top-down tree, but each node also has a horizontal bar figure representing the genetic part's DNA sequence. This figure will indicate what partitions of the DNA sequence consist of its children's DNA sequences. We decided to add this DNA sequence figure because Michael said a visual representation of how the sequence dependencies manifest would be useful and isn't available anywhere currently. We thought this would make our visualization more unique and much more useful.



Must-Have Features.

- Search box with the ability to click on a result and render a tree.
- Tree rendering with minimal tooltip to tell you the name of a node.
- Full details in detail view on click.

Optional Features.

- SBOL glyphs as nodes in tree.
- DNA sequence visualization.
- Tree expand/collapse.
- Search relevance horizontal bars.
- Improved tooltip on hover.

Project Schedule:

Week of Oct. 29: Project Setup, Data Gathering

Week of Nov. 5: Search, Basic Tree Display,

Week of Nov. 12: Hover and click boxes, tree sequence figures

Week of Nov. 19: Polish Tree, Tree Collapsing

Week of Nov. 26: (Nov. 30 due) Buffer