

# Force Direct Visualization for Phylogenetic Trees

Projeto e Seminário

Licenciatura em Engenharia Informática e Computadores

Vasco Branco Revés  
a44818@alunos.isel.pt  
96611441

Supervisors:

Cátia Vaz, cvaz@cc.isel.ipl.pt, ISEL  
Ana Correia, ana.duarte.correia@ist.utl.pt, IST

April 12, 2021

## 1 Introduction

Nowadays more than ever, epidemics have become an issue of increasing importance, the study of biologic sequences and epidemiological data is essential and seeing an exponential growth in the world.

Epidemiological surveillance is now a global procedure rather than a country-based one. Combining information of country-specific data-sets can now reveal epidemic spreading patterns that were not possible to detect before, but phylogenetic tree visualization algorithms are often hard to use and integrate in analysis frameworks and tools.

Phyloviz (1) is an open source platform that provides analysis of sequence-based typing methods that generate allelic profiles.

DNA sequencing is the process used to determine the nucleic acid sequence in a DNA molecule. Sequence-based typing methods include any method or technology used to determine the order of the four DNA bases: adenine, guanine, cytosine and thymine (2).

Choosing the typing method to use depends on the context that is being used and the problem to solve. Most recent technologies allow high speed DNA and RNA sequencing, called High Throughput Sequencing.

These techniques allow the characterization of bacteria at the strain level providing researchers with important information for the surveillance of infectious diseases, outbreak

investigation, pathogenesis studies, natural history of infection and bacterial population genetics (3).

With the information obtained from these techniques it's possible to estimate a Phylogenetic Tree, which is a representation of the relationships of gene or protein sequences to their ancestral sequences and complementary information.

The study of Phylogenetic trees is used to differentiate microorganisms at the sub-species or strain level, allowing to understand the evolutionary history of gene families, tracing the origin and transmission of infectious diseases (4).

Phyloviz is available as Desktop application developed in Java, available for all platforms, and online application, Phyloviz-Online (5).

The goal of this project is to develop a library for phylogenetic tree visualization, allowing large amounts of data when building phylogenetic trees without damaging the performance and to improve the user interface, this includes responsive graphics, statistics, filters, etc.

## 2 Requirements

Our aim with this project is to provide a solution to be included in Phyloviz-Online platform that uses a Force Direct Layout (6) for phylogenetic trees visualization.

This solution will be independent from the other modules and it will be developed in Javascript.

The server side application is still being developed and the best platform to develop the application are still to be chosen. This decision will be based on the compatibility with other libraries, community help and documentation. The possibilities so far are NWjs<sup>1</sup>, React<sup>3</sup>, Electron<sup>2</sup>.

The project will consist on the following requirements:

1. **Force direct layout** for Graph visualization in the browser. This algorithm is based on a physical model, simulating physical forces among the set of edges and nodes, based on their relative positions and using this forces to simulate movement. The nodes are represented by points that repel each other like magnets resulting in as few crossing edges as possible. The Layout of the graph is based solely on the information from the graph data. For this we're planning to use the Sigma js library but more tests will be done in order to find out the best libraries to use in graphic drawing and implement the algorithm.

2. **Collapse & Expand**, to allow user to collapse or expand a specific region of the graph.

3. **Pie-Chart Graphics**. Includes the possibility to add or remove pie-chart graphics to the nodes of the tree.

---

<sup>1</sup>NWjs - <https://nwjs.io/>

<sup>3</sup>ReactJs - <https://reactjs.org/>

<sup>2</sup>Electron - <https://www.electronjs.org/>

4. **Labels & Filters.** The application will include the option to add labels to the nodes or connections between nodes, making it easier to read the graph. Also, it should include the possibility to search and filter the data, for example, search by country, sex, age, etc.

5. **Statistics.** The application should include simple statistics, to better comprehend the impact or the relevance of the information the user is reading. For example percentage per country or number of isolates associated to a strain.

6. **Save Image State.** It should be able to save the state of visualization, this is, the tree, positions, associated filters, etc.

### 3 Schedule

The project's schedule is described in table 1:

Start Date	Duration (in weeks)	Description
11/03/21	2	- Introduction to Phyloviz and Phyloviz-Online Platform - Preliminary Studies concerning Biology concepts
25/03/21	2	- Writing project proposal - Study of graphic design platforms available (d3, Sigma, VivaGraph) - Study of pie-chart design techniques/algorithms
06/04/21	1	- Finish and deliver Project Proposal
12/04/21	2	- Testing graph design platforms performance (Speed, compatibility) - Testing platforms performance
26/04/21	2	- Study API's architecture - Start implementing force directed algorithms. - Pie-chart graphics algorithms - How to add filters/labels
10/05/21	2	-Start implementing algorithms in the phyloviz-Online platform -Writing Progress Report
24/05/21	2	-Deliver Progress Report -Continue implementation - Start Tests - How to save state
07/06/21	2	- Finish Tests and refining last details

Table 1: Schedule.

## References

- [1] Marta Nascimento, Adriano Sousa, Mário Ramirez, Alexandre P Francisco, João A Carriço, and Cátia Vaz. Phyloviz 2.0: providing scalable data integration and visualization for multiple phylogenetic inference methods. *Bioinformatics*, 33(1):128–129, 2017.
- [2] Michael M. Miyamoto and Joel Cracraft. *Phylogenetic Analysis of DNA Sequences*. Oxford University Press, 1991.
- [3] Alexandre P Francisco, Cátia Vaz, Pedro T Monteiro, José Melo-Cristino, Mário Ramirez, and João A Carriço. Phyloviz: phylogenetic inference and data visualization for sequence based typing methods. *BMC Bioinformatics*, 13:87, 2012.
- [4] João A. Carriço, Maxime Crochemore, Alexandre P. Francisco, Solon P. Pissis, Bruno Ribeiro-Gonçalves, and Cátia Vaz. Fast phylogenetic inference from typing data. *Algorithms for Mol Bio.*, 2018.
- [5] Bruno Ribeiro-Gonçalves, Alexandre P Francisco, Cátia Vaz, Mário Ramirez, and João André Carriço. Phyloviz online: web-based tool for visualization, phylogenetic inference, analysis and sharing of minimum spanning trees. *Nucleic acids research*, 44(W1):W246–W251, 2016.
- [6] Se-Hang Cheong and Yain-Whar Si. Force-directed algorithms for schematic drawings and placement: A survey. *Information Visualization*, 19(1):65–91, 2020.