

PHYLOViZ Online: Web-based Phylogenetic Data Analysis And Visualization For Allelic Profiles And SNP Data

Bruno Gonçalves¹, Alexandre Francisco^{2,3}, Cátia Vaz^{2,4}, Mário Ramirez¹, João Carriço¹

(1) Instituto de Microbiologia, Instituto de Medicina Molecular, Faculdade de Medicina, Universidade de Lisboa, Lisboa, Portugal

(2) INESC-ID, R. Alves Redol 9, 1000-029 Lisboa, Portugal

(3) Instituto Superior Técnico, Universidade de Lisboa, Lisboa, Portugal

(4) Instituto Superior de Engenharia de Lisboa, Instituto Politécnico de Lisboa, 007 Lisboa, Portugal

Introduction

PHYLOViZ Online is a user-friendly web application, an online version of the standalone PHYLOViZ(1), developed to allow users to perform profile based data analyses without software installation, and to enable an easy data access, analysis, and sharing from any Internet enabled computer. It allows data set storage and visualization through distinct visual strategies, and a programmatic access to data through a RESTful API allowing it to be seamlessly integrated into any third party web service or software.

PHYLOViZ Online is freely available at <https://online.phyloviz.net>

User Specific Area

- **STORE** your own data sets.
- **SHARE** them through an *user specific URL*.
- Or by making your data sets **PUBLICLY** available.

Scan to see the
Video Tutorial!



Modify the Minimum Spanning Tree

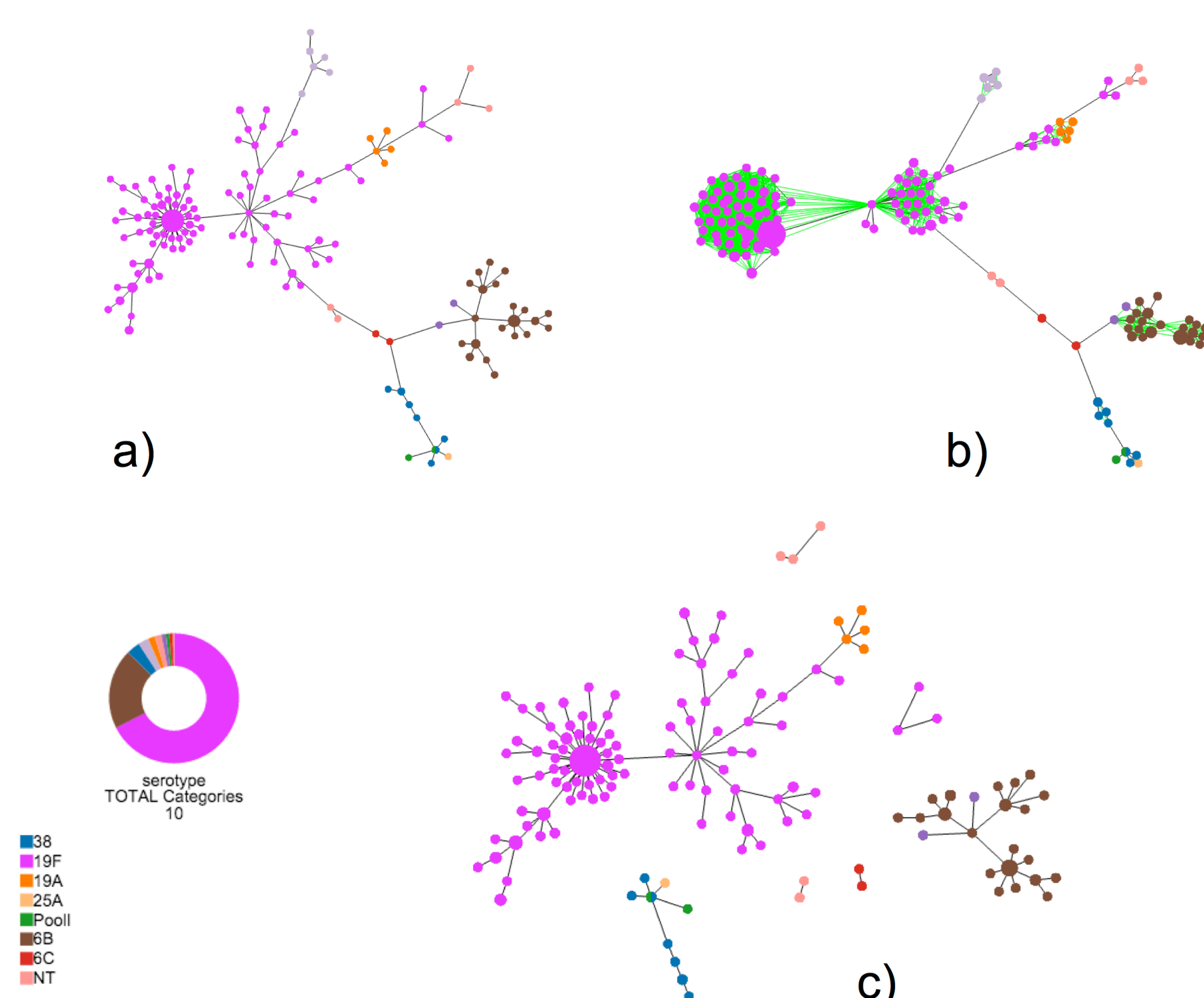


Figure 2: Different features available that modify the Minimum Spanning Tree. a) Colour assignment. Colours assigned by serotype field existing in an auxiliary data file. b) n Locus Variant. All nodes with distances equal or above 12 are linked. c) Tree cut-off. All links with a distance value above 60 were deleted.

- **Attribute colours** to nodes according to auxiliary data fields.
- **LINK** all nodes at a given distance (*nLV graph*)
- **HIDE** all links above a given distance threshold (*Tree cut-off*).

RESTful API

Application Programming Interface for increased interoperability:

- **RETRIEVE** programmatically public or registered data.
- **UPLOAD** data.
- **RUN** available tree algorithms on stored data.

More information at: <https://online.phyloviz.net/api>

Conclusions

PHYLOViZ Online is a freely available application that provides an effective way for users to visualize, perform visual analytics and share annotated data sets. It also complies with the increasing requirement to make complete data sets available and reproducible so that they can be independently scrutinized. PHYLOViZ Online source code is freely available under GPLv3 licence at https://github.com/bfgoncalves/Online-PhyloViZ/tree/postgres_version. A set of Node.js modules is also available at https://www.npm.com/package/phyloviz_bundle for developers to incorporate PHYLOViZ tree visualization into their software.

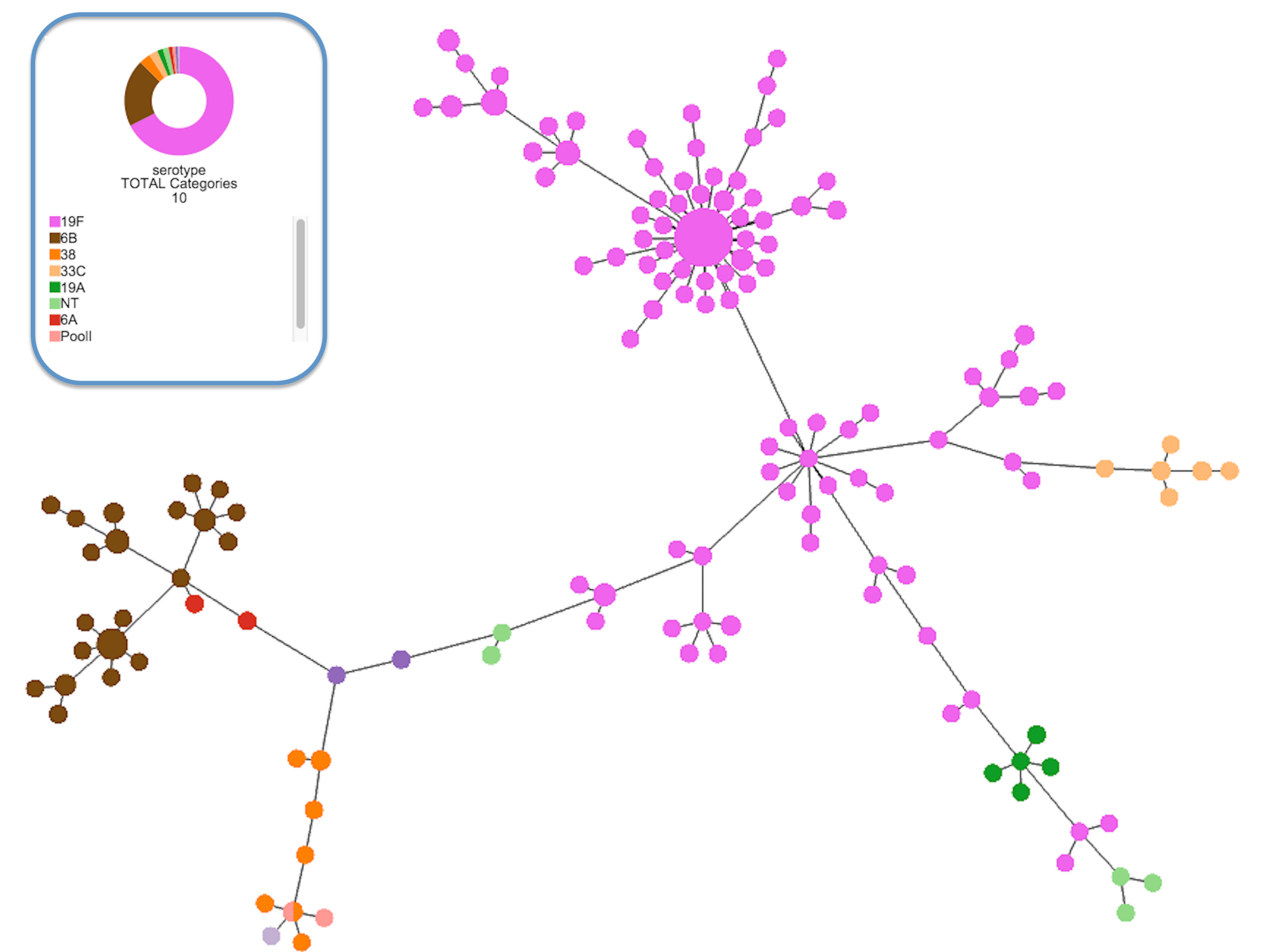


Figure 1: PHYLOViZ Online Tree visualization of a data set that consists of 145 strains of *Streptococcus pneumoniae* whose partial genome sequences are publicly available(2, 3). The profile data consists on a 237 loci cgMLST scheme. The Minimum Spanning Tree was computed using the goeBURST algorithm(2) and Colours represent distinct serotypes provided using an auxiliary data file.

Interactive Distance Matrix

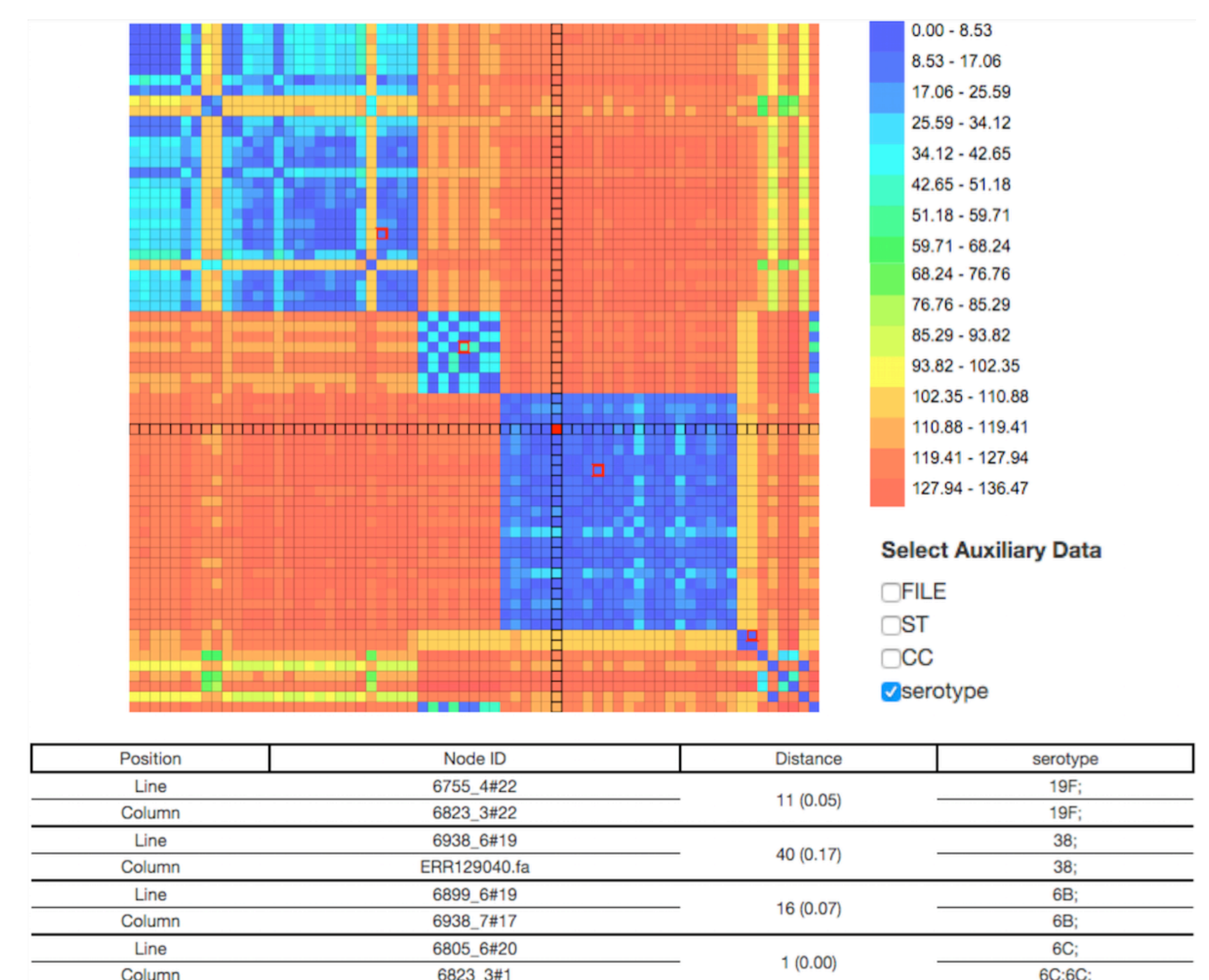


Figure 3: Interactive Distance Matrix constructed from a node selection of a cgMLST *Streptococcus pneumoniae* data set. Matrix cells were ordered according to serotype and colours represent distances computed through pairwise comparisons of profiles. Selected nodes on the distance matrix (red border colour) have their information displayed in the table according to the selected auxiliary data fields.

- a) **VISUALIZE** pairwise distances of profiles in a heatmap display.
- b) **ORDER** cells according to auxiliary data fields.
- c) **EXPLORE** the distance matrix using a dynamic table.

References

1. Francisco,A.P., Vaz,C., Monteiro,P.T., Melo-Cristino,J., Ramirez,M. and Carriço,J.A. (2012) PHYLOViZ: phylogenetic inference and data visualization for sequence based typing methods. *BMC Bioinformatics* 2015 16:1, 13, 87.
2. Chewapreecha,C., Harris,S.R., Croucher,N.J., Turner,C., Marttinen,P., Cheng,L., Pessia,A., Aanensen,D.M., Mather,A.E., Page,A.J., et al. (2014) Dense genomic sampling identifies highways of pneumococcal recombination. *Nature genetics*, 46, 305–309.
3. Croucher,N.J., Finkelstein,J.A., Pelton,S.I., Mitchell,P.K., Lee,G.M., Parkhill,J., Bentley,S.D., Hanage,W.P. and Lipsitch,M. (2013) Population genomics of post-vaccine changes in pneumococcal epidemiology. *Nature genetics*, 45, 656–663.
4. Francisco,A.P., Bugalho,M., Ramirez,M. and Carriço,J.A. (2009) Global optimal eBURST analysis of multilocus typing data using a graphic matroid approach. *BMC Bioinformatics* 2015 16:1, 10, 152.

Acknowledgments

BRG was supported by grant SFRH/BD/101448/2014 from the Fundação para a Ciência e Tecnologia, Portugal.