OXFORD

Genetics and population analysis

tcsBU: a tool to extend TCS network layout and visualization

António Múrias dos Santos[1,2,*], Maria Pilar Cabezas[2], Ana Isabel Tavares[1,2], Raquel Xavier[2,3] and Madalena Branco[2]

[1]Faculdade de Ciências da Universidade do Porto, 4169-007 Porto, Portugal [2]CIBIO, Centro de Investigação em Biodiversidade e Recursos Genéticos, Campus Agrário de Vairão, 4485-661 Vairão, Portugal and [3]School of Biological Sciences, Cardiff University, Cardiff, CF10 3AX, UK

*To whom correspondence should be addressed.
Associate Editor: Janet Kelso

Abstract

Motivation: TCS is a widely used haplotype network reconstruction software, but lacks the capability of overlapping genetic with geographic structure, which is often a first step in phylogeographic analysis.

Results: tcsBU is a web-based program that extends the capabilities of TCS, by implementing haplotype classification into an arbitrary user-defined scheme, which is displayed as pie-chart like graphs embedded into the network. Taking advantage of modern graphic libraries, tcsBU also improves the speed at which the final network layout is reached. Networks can be saved as a Scalable Vector Graphics format.

Availability and implementation: tcsBU is available on-line at http://cibio.up.pt/software/tcsBU/. The source code is freely available from https://github.com/sairum/tcsbu/ under a standard MIT license.

Contact: amsantos@fc.up.pt

1 Introduction

Construction of haplotype networks that represent genealogical relationships among genes at the population level is often a first step of any phylogeographic analysis. Because gene-genealogies are usually polytomous and network-like, traditional tree-building methods, which force dichotomous branching patterns, are seldom applicable (Bandelt *et al.*, 1995). Different methods were developed to overcome such problems (e.g. Bandelt *et al.*, 1999; Excoffier and Lischer, 2010). Among these, TCS (Clement *et al.*, 2000) is unique, in that it implements the probabilistic method of *Statistical Parsimony* (Templeton *et al.*, 1992). This allows users to arbitrarily define the probability at which haplotypes connect with each other, which may yield multiple networks per analysis. This feature has been suggested as a species delimitation tool in barcoding analysis (Hart and Sunday, 2007).

Although the core TCS algorithm (network computation) is fast, there are some major limitations related with its graphical output. In most phylogeographic analyses, it is often useful to combine genetic structure (network topology) with geographic structure (haplotype occurrence), but TCS has no built-in way of classifying haplotypes and displaying that information in the final network. TCS cannot use its 'spring' method to layout multiple networks, meaning that the resulting overlapped networks should to be disentangled manually using an external vector manipulation software. Moreover, the 'spring' method is extremely slow to layout large or complex networks. As a consequence, it takes a significant effort to edit the resulting TCS vector file to produce publication-ready networks.

2 Overview

tcsBU was developed to extend TCS capabilities, namely by providing a way to arbitrarily classify haplotypes and to display that information on the final network. Its force-directed layout algorithm is much faster than TCS's own 'spring' method, and can handle multiple network results easily (Fig. 1). Using any modern browser, users are exposed to an
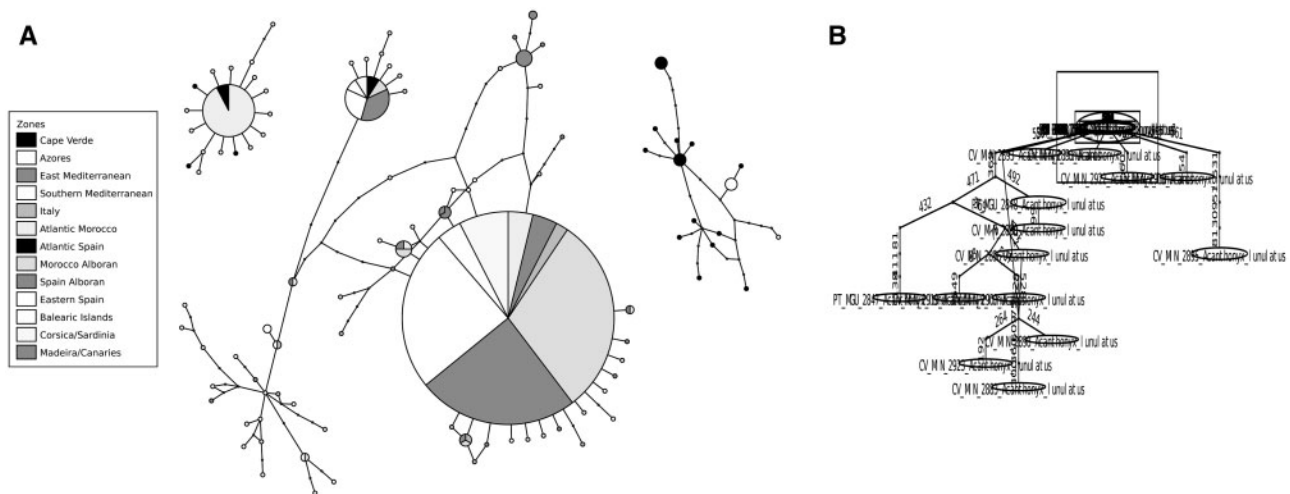
Fig. 1. (A) multiple networks and legend made in tcsBU and depicted without any additional editing; **(B)** the same networks exported from TCS

intuitive user-interface through which they can edit the network and save it as a Scalable Vector Graphics (SVG) format, a World Wide Web Consortium standard. The SVG file can be directly embedded in some word processors, or edited and exported to other graphic formats using a vector manipulation program.

tcsBU is a browser-based javascript program, that can be served by an HTTP server or can be installed into a local folder and run by any modern browser. An on-line version is available from http://cibio.up.pt/software/tcsBU/. tcsBU does not impose any burden on the server besides serving the bundled javascript libraries. All computations are done by the browser itself. The software depends on a few third-party,freely available libraries: *JQuery* (https://jquery.com), *w2ui* – JQuery-based UI (http://w2ui.com), *d3.js* – force-directed layout engine (https://d3js.org), *FileSaver.js* – implements file saving in the browser (https://github.com/eligrey/FileSaver.js) and *d3-svg-legend* (https://github.com/emeeks/d3-svg-legend). Because many of these libraries dropped support for old versions of some browsers, tcsBU will only run on modern ones (Firefox 20+, Chrome, IE 10+, Opera 15+ and Safari 6.1+).

2.1 Installation

For a web-based installation, the tcsBU package (download from https://github.com/sairum/tcsbu/) can be saved into any directory under the web-server ROOT. No further steps are necessary, and users will be able to access the software by pointing their browsers at the specified location. For a local installation, tcsBU can be saved into any directory and run using the *file://* protocol. If a network connection is available, users can click on the *index.html* file, which should automatically open the default browser. Otherwise, users should use *tcsbu.html*. The difference is that the latter is an off-line version which will not download the necessary libraries (JQuery, w2ui, d3.js) from their respective Content Delivery Network sites, but will use the ones shipped with the program.

2.2 Usage

The first step is to load a TCS output file. There is no need to use TCS post-analysis facilities, such as the 'spring' algorithm, to disentangle the network once the network topology is computed. The force-directed layout algorithm of tcsBU is much faster, and can cope with multiple network scenarios. Users should look for a Graph Modeling Language (GML) file with extension.*graph*, usually found on the directory where

the alignment used by TCS is located. The user interface is self-intuitive, and there is an Help window with extensive information on how to install the software, refine the parameters of the force-directed layout algorithm, or perform haplotype classification. Example files are also provided in the UI. Users can stop and restart the engine at any time, and they can drag (reposition) nodes during run-time.

3 Conclusion

tcsBU extends TCS capabilities, and speeds up the production of publication-ready networks resulting from TCS analysis. Source code is freely available from GitHub and improvements/bug fixes are welcome.

Funding

Conflict of Interest: none declared.

References

Bandelt,H.-J. *et al.* (1995) Mitochondrial portraits of human populations using median networks. *Genetics*, **141**, 743–753.

Bandelt,H.J. *et al.* (1999) Median-joining networks for inferring intraspecific phylogenies. *Mol. Biol. Evol.*, **16**, 37–48.

Clement,M. *et al.* (2000) TCS: a computer program to estimate gene genealogies. *Mol. Ecol.*, **9**, 1657–1659.

Excoffier,L. and Lischer,H.E.L. (2010) Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Mol. Ecol. Resources*, **10**, 564–567.

Hart,M.W. and Sunday,J. (2007) Things fall apart: biological species form unconnected parsimony networks. *Biol. Lett.*, **3**, 509–512.

Templeton,A. *et al.* (1992). A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping and DNA sequence data. iii. Cladogram estimation. *Genetics*, **132**, 619–633.