bioRxiv preprint doi: https://doi.org/10.1101/169177; this version posted October 13, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

BIOINFORMATICS

Vol. 00 no. 00 2017
Pages 1–3

SPECTRE: a Suite of PhylogEnetiC Tools for Reticulate Evolution

Sarah Bastkowski [1*†], Daniel Mapleson [1†], Andreas Spillner[2], Taoyang Wu [3], Monika Balvočiūtė[4] and Vincent Moulton [3]

[1]Earlham Institute, Norwich Research Park, Norwich, NR4 7UZ, UK; [2]Merseburg University of Applied Sciences, 06217 Merseburg, Germany; [3]School of Computing Sciences, University of East Anglia, Norwich, NR4 7TJ, UK; [4]University of Tübingen, Department of Computer Science, Sand 14, 72076, Tübingen, Germany;

Received on XXXXX; revised on XXXXX; accepted on XXXXX

Associate Editor: XXXXXXX

ABSTRACT

Summary Split-networks are a generalization of phylogenetic trees that have proven to be a powerful tool in phylogenetics. Various ways have been developed for computing such networks, including split-decomposition, NeighborNet, QNet and FlatNJ. Some of these approaches are implemented in the user-friendly SplitsTree software package. However, to give the user the option to adjust and extend these approaches and to facilitate their integration into analysis pipelines, there is a need for robust, open-source implementations of associated data structures and algorithms. Here we present SPECTRE, a readily available, open-source library of data structures written in Java, that comes complete with new implementations of several pre-published algorithms and a basic interactive graphical interface for visualizing planar split networks. SPECTRE also supports the use of longer running algorithms by providing command line interfaces, which can be executed on servers or in High Performance Computing (HPC) environments.

Availability: Full source code is available under the GPLv3 license at: `https://github.com/maplesond/SPECTRE`
SPECTRE's core library is available from Maven Central at: `https://mvnrepository.com/artifact uk.ac.uea.cmp.spectre/core`
Documentation is available at: `http://spectre-suite-of-phylogenetic-tools-for-reticulate-evolution.readthedocs.io/en/latest/`

Contact: sarah.bastkowski@earlham.ac.uk

Supplementary Information (SI): Supplementary information is available at Bioinformatics online.

1 INTRODUCTION

Split-networks are a generalization of phylogenetic trees that are commonly used to analyze reticulate evolution in organisms such as plants, bacteria and viruses (see Figure 1 for an example). They provide a snapshot of the data and can be used to display conflicting signals. Examples of algorithms for computing such networks include split-decomposition (Bandelt and Dress, 1992), Neighbor-Net (Bryant and Moulton, 2004), QNet (Grünewald *et al.*, 2007), SuperQ (Grünewald *et al.*, 2013) and FlatNJ (Balvočiūtė *et al.*, 2014). A comprehensive overview of split-networks can be found in (Huson and Bryant, 2006).

Currently the main program available for computing split-networks is the user-friendly SplitsTree program (Huson and Bryant, 2006). In addition, various methods for computing split-networks such as some of those mentioned above have been implemented and released as stand alone applications. Implementing data structures capable of representing the mathematical structures used to describe and compute split networks is not a trivial undertaking and existing software either is closed source or have their data structures and algorithms tightly integrated with their host tool, so are not easily reusable. There are therefore currently few options for developers wishing to create or extend their own tools based on these concepts other than to start from scratch. Hence there is a need for a robust and flexible open-source library that provides core data structures and algorithms to facilitate development of new tools and algorithms.

2 SPECTRE

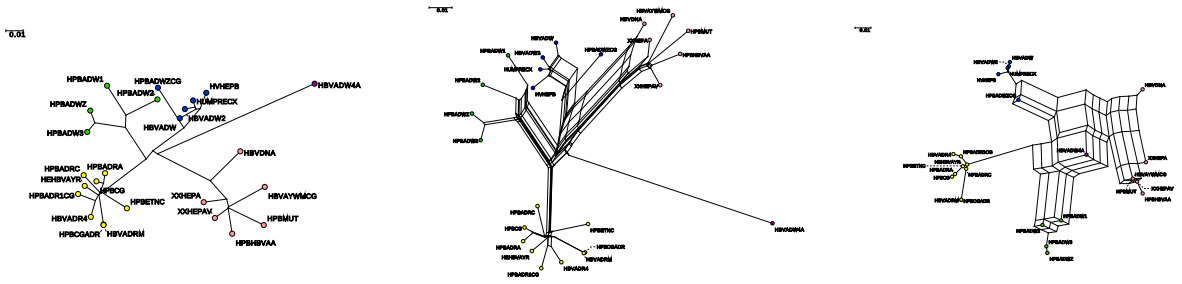
Here we present SPECTRE, a suite of tools for computing, modelling, and visualizing reticulate evolution based on split-networks. SPECTRE builds in part on existing open-source implementations of some of these tools, in particular for QNet, SuperQ and FlatNJ, integrating them into a unified and extendible library. The main tools available through SPECTRE are summarized below (for more details see section 1 of SI):

- NeighborNet rapidly constructs a circular split network from a distance matrix or a sequence alignment (Bryant and Moulton, 2004). NetMake implements variants of NeighborNet as described in (Levy and Pachter, 2011).
- SuperQ constructs a circular split network from a set of (partial) input trees (Grünewald *et al.*, 2013).

*to whom correspondence should be addressed
†equal contribution

© Oxford University Press 2017.

bioRxiv preprint doi: https://doi.org/10.1101/169177; this version posted October 13, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.



(a) Minimum Evolution tree (b) NeighborNet network (c) FlatNJ network

Fig. 1: To illustrate some of SPECTREs functionality, we processed a data set analyzed in (Bollyky *et al.*, 1996) consisting of different Hepatitis B viruses (HBV). There are 5 different genomic groups and the phylogenetic analysis led to the result that HBVDNA is a recombinant with around half the genome coming from group A and half from group D. It also concluded that HPBADW1 is a recombinant of HPBADW2 (B) and HPBADWZCG (A), but with only a small insertion from HPBADWZCG into the Genome. (a) shows a minimum evolution tree constructed by NetME that is compatible with the split network constructed by NeighborNet, which is shown in (b). (c) shows the split network constructed by FlatNJ.

- FlatNJ constructs a flat split network from a multiple sequence alignment, weighted quartet data or location data (Balvočiūtė *et al.*, 2014).
- NetME produces a minimum evolution tree compatible with an existing circular split network (Bastkowski *et al.*, 2014).

These tools are accessible to the user via graphical and command line interfaces. Apart from driving the tools, the interactive graphical interface can visualize planar split networks using the drawing algorithm in (Spillner *et al.*, 2012). The interface offers a number of basic functions for orientating the canvas (e.g. zoom, pan, flip and rotate), manipulating labels (size, color, location) and creating image files (PDF, EPS, SVG, PNG). The command line implementation enables bioinformaticians to integrate tools into pipelines. This works on desktop PCs, like SplitsTree, but is also designed so long running tools are executable on servers or high performance computing environments where displays are not available.

For developers wishing to reuse code and develop their own tools, SPECTRE provides a core library containing common data structures (e.g. splits, trees, networks, distances, quartets and multiple sequence alignments), algorithms (e.g. NeighborNet), and robust file parsers to process a range of input files (e.g. NEXUS, PHYLIP, Newick, Emboss, FastA); see section 2 of SI for more details. The library is available directly from Maven Central, giving developers direct access to the most recent version of the library and and providing a convenient way to integrate it into the processes for building their own projects.

3 CONCLUDING REMARKS

SPECTRE provides a collection of open-source tools and resources for modelling, understanding and visualizing reticulate evolution based on split networks. We believe that our software will both enable bioinformaticians to easily test and compare methods for inferring planar split networks and help computer scientists build their own methods for inferring phylogenetic networks by reusing our existing data structures and algorithms via the open-source library. Moreover, this also provides the option to easily add such new tools to the library making them readily available to other users.

ACKNOWLEDGMENTS

We would like to thank Stephan Grünewald.

REFERENCES

Balvočiūtė, M., Spillner, A. and Moulton, V. (2014). FlatNJ: A novel network-based approach to visualize evolutionary and biogeographical relationships. *Systematic Biology*, **63**(3), 383–396.

Bandelt, H.J. and Dress, A. (1992). Split decomposition: a new and useful approach to phylogenetic analysis of distance data. *Molecular Phylogenetics and Evolution*, **1**, 242–252.

Bastkowski, S., Spillner, A. and Moulton, V. (2014). Fishing for minimum evolution trees with NeighborNets. *Information Processing Letters*, **114**(1), 13–18.

Bollyky, P.L. et al (1996). Recombination between sequences of hepatitis B virus from different genotypes. *Journal of Molecular Evolution*, **42**, 97–102.

Bryant, D. and Moulton, V. (2004). Neighbor-Net: An agglomerative method for the construction of phylogenetic networks. *Molecular Biology and Evolution*, **21(2)**, 255–265.

Grünewald, S. et al (2007). QNet: An agglomerative method for the construction of phylogenetic networks from weighted quartets. *Molecular Biology and Evolution*, **24**, 532–538.

Grünewald, S. et al (2013). SuperQ: Computing supernetworks from quartets. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, **10**(1), 151–160.

Huson, D. and Bryant, D. (2006). Application of phylogenetic networks in evolutionary studies. *Molecular Biology and Evolution*, **23**, 254–267.

Levy, A. and Pachter, L. (2011). The Neighbor-Net algorithm. *Advances in Applied Mathemathics*, **47**(2), 240–258.

bioRxiv preprint doi: https://doi.org/10.1101/169177; this version posted October 13, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Spillner, A., Nguyen, B.T. and Moulton, V. (2012). Constructing and drawing regular planar split networks. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, **9(2)**, 395–407.