SPRINT: Package for Calculating and Visualising the Minimal Hybrid Number of a Phylogenetic Network Manual

Liam J Maher, Taoyang Wu, Katharina T Huber

25 August, 2022

1 Introduction

Overview Polyploidy is an important evolutionary driver which has been observed in plants, bacteria, animals and more. An existing algorithm for reconstructing the evolutionary history of polyploid species can be found in the paper K. T. Huber *et al* (2006), however, this method relies on the existence of so called multilabelled trees which may not be readily available as obtaining these is not trivial. There is currently no tool available which computes the minimum number of hybridisation vertices possible to explain a given set of polyploid species. Here, we present the Species Ploidy Realisation of Integers with Networks Tool (SPRINT) software package that has been developed to circumvent the reliance on multilablled trees when obtaining an evolutionary picture from a polyploid dataset.

Installation The program SPRINT is freely available from **ENTER URL HERE**. This page describes how to install the program and offers a number of test files to run as examples. The only requirements are that the machine on which the software is being run has Python virtual machine 3.9 or later installed and it is capable of displaying a graphical user interface. Any queries about the installation of the program should be directed to l.maher@uea.ac.uk.

Running the program

Disclaimer This software is supplied as-is, with no warranty of any kind expressed or implied. We have made every effort to avoid errors in design

and execution of this software, but we will not be liable for its use or misuse. The user is solely responsible for the validity and consequences of any results generated.

2 Using the program

Data entry window When the program first loads you will be presented with a data entry window within which are a number of options to enter data and how users would like to visualise data, Figure 1.

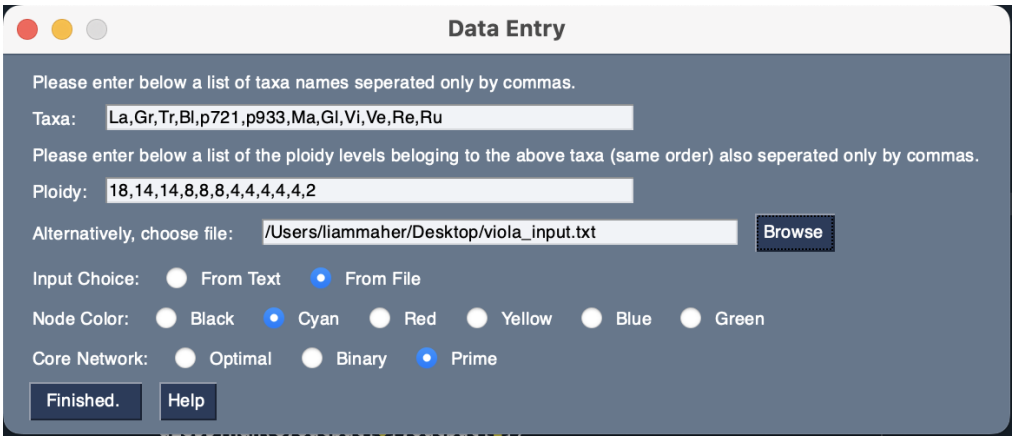

Figure 1: Data entry window.

The view of this window may vary slightly depending on the operating system and display resolution of an individual user.

Text entry The first option to enter data into the program is composed of two input text boxes. One is for the species taxa and the other for those species' ploidy levels. For taxa, users input alphanumeric strings, one for each taxa of interest, separated by a comma (no spaces). For ploidy levels, users input integers greater than 0 for each taxa entered above, in the same order as entered above and separated by a comma (no spaces).

File browser The second option to enter data into the program is by searching on your system for a .txt file via the 'Browse' button found within the data entry window. The .txt file should be of the form:

- **First line** For taxa, users input alphanumeric strings, one for each taxa of interest, separated by a comma (no spaces).

- **Second line** For ploidy levels, users input integers greater than 0 for each taxa entered above, in the same order as entered above and separated by a comma (no spaces).

Input choice Once the user has either input data into the data entry text boxes or alternatively the user has browsed for and selected a .txt file containing the data and therefore the file path has been displayed, the user then selects which method they have chosen from the radio buttons labelled 'Input Choice:' by selecting 'From Text' or 'From File', respectively.

Note: selecting the 'From File' radio button before the file path has been displayed will cause the program to fail and result in an error message.

Node color The radio buttons beside the label 'Node Color:' gives the user six different color options. When the program produces the phylogenetic network visualisation of the input data, the choice of color made by the user will be reflected in the color of the nodes/vertices in the output figure. The default option is 'Black' but any color can be chosen with no effect on the run time of the program.

Core network The radio buttons beside the label 'Core Network:' gives the user three different options from which to initialize the programs algorithm. Radio buttons 'Binary' and 'Prime' refers to the core network construction methods Binary representation and Prime factor decomposition respectively, see (Huber and Maher, 2022) for more information on core network construction methods. The default option, 'Optimal', will compare the number of hybrid vertices between two different phylogenetic networks constructed with the core network construction methods listed above. The program will then choose the network with the fewest hybrids to initialize the remainder of the construction.

Output Once the data has been entered and the user has selected their preferences in the Data entry window, click the 'Finished' button to proceed. For example, if the .txt file found in the documents of this program was used then the ouput would look similar to Figure 2.

The title of the output GUI window displays the ploidy profile used as input. The top of the window contains a figure which displays a phylogenetic network. This network realises the input ploidy profile and contains the fewest possible number of hybrid vertices given the input ploidy profile.

Shown below the figure is the simplification sequence, starting from the input ploidy profile at the top and continuing until the final element is a simple ploidy profile, from which the core network is constructed and the

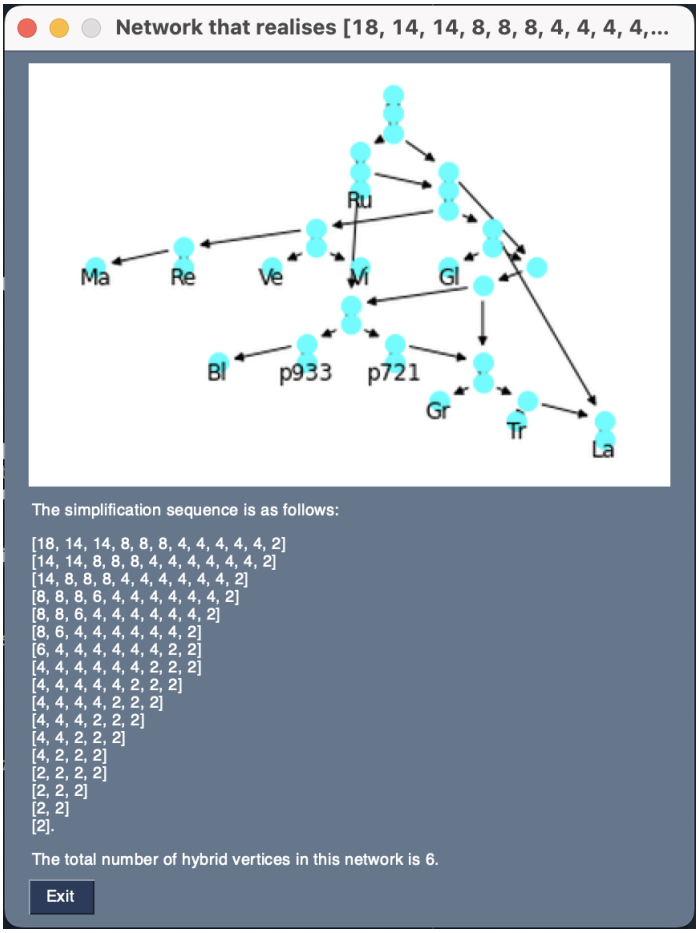

Figure 2: Output window.

traceback begins.

Finally, at the bottom of the window, the minimal number of hybrid vertices for the input ploidy profile is displayed. In Figure 2, the minimal hybrid number is 6.

3 Legal

Whilst this implementation of the algorithm presented in K. T. Huber and L. J. Maher (2022) is complete we still continue to actively work on SPRINT. SPRINT will always remain free of charge however, during this development process the source code available on the website may not be the most recent version.

4 References

K. T. Huber and L. J. Maher (2022). The Hybrid Number of a Ploidy Profile, ArXiv.