

Phylogenetics and Molecular Evolution/Filogenética e Evolução Molecular

Octávio S. Paulo

Computational Biology and Population Genomics Group (CoBiG2)

Análise intraespecífica de sequências

Sumário:

Redes intraespecíficas.



Ciências
ULisboa

Faculdade
de Ciências
da Universidade
de Lisboa

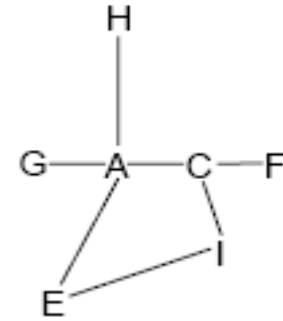


Computational
Biology & Population
Genomics Group

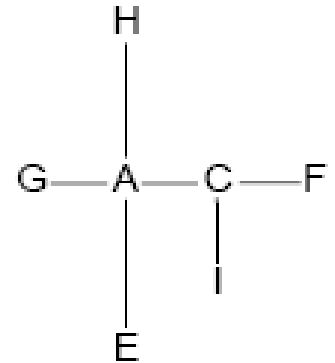


Networks

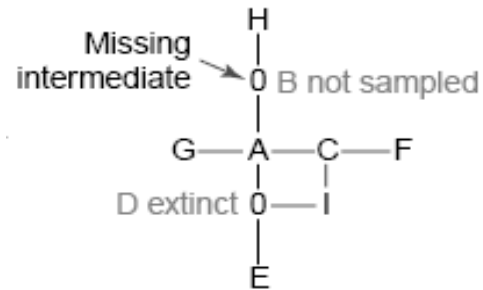
1. Minimum spanning network



2. Median Joining Network



3. Statistical Parsimony



Minimum-spanning network

An algorithm for constructing minimum-spanning trees (MSTs) from a matrix of pairwise distances (absolute number of differences) among haplotypes (Prim, 1957; Rohlf, 1973) has been modified in order to include all possible MSTs within a single graph, the MSN (Excoffier and Smouse, 1994). The connections in MSTs are only formed among sampled haplotypes. The inference of missing node haplotypes is therefore not possible. Software ARLEQUIN, v. 2.000 (Schneider et al., 2000).

Median Joining Networks

Bandelt H-J, Forster P, Sykes BC, Richards MB (1995)
Mitochondrial portraits of human populations using median
networks. Genetics 141:743-753

Bandelt, Hans-Jurgen, et al. Median-joining networks for
inferring intraspecific phylogenies. Molec Bio and Evol. Vol 16
(1), Jan 1999

Median-joining network

Under this approach (Bandelt et al., 1999), all MSTs are first combined within a single network (MSN) following an algorithm analogous to that proposed by Excoffier and Smouse (1994). Then, using the parsimony criterion, inferred missing node haplotypes are added to the graph in order to reduce its overall length. Program NETWORK, v. 2.0 (available at <http://www.fluxus-engineering.com/sharenet.htm>), with parameter $\epsilon = 0$. Additionally, some of the data sets were reanalyzed with an increased value of epsilon to test whether it improves the performances of the method (with $\epsilon > 0$, less parsimonious pathways are also included in the graph).

Median-joining networks for inferring intraspecific phylogenies

H J Bandelt ¹, P Forster, A Röhl

Affiliations + expand

PMID: 10331250 DOI: [10.1093/oxfordjournals.molbev.a026036](#)

Abstract

Reconstructing phylogenies from intraspecific data (such as human mitochondrial DNA variation) is often a challenging task because of large sample sizes and small genetic distances between individuals. The resulting multitude of plausible trees is best expressed by a network which displays alternative potential evolutionary paths in the form of cycles. We present a method ("median joining" [MJ]) for constructing networks from recombination-free population data that combines features of Kruskal's algorithm for finding minimum spanning trees by favoring short connections, and Farris's maximum-parsimony (MP) heuristic algorithm, which sequentially adds new vertices called "median vectors", except that our MJ method does not resolve ties. The MJ method is hence closely related to the earlier approach of Foulds, Hendy, and Penny for estimating MP trees but can be adjusted to the level of homoplasy by setting a parameter epsilon. Unlike our earlier reduced median (RM) network method, MJ is applicable to multistate characters (e.g., amino acid sequences). An additional feature is the speed of the implemented algorithm: a sample of 800 worldwide mtDNA hypervariable segment I sequences requires less than 3 h on a Pentium 120 PC. The MJ method is demonstrated on a Tibetan mitochondrial DNA RFLP data set.

Free Phylogenetic Network Software

Network generates evolutionary trees and networks from genetic, linguistic, and other data. Network can then provide age estimates for any ancestor in the tree.

Disclaimer: Network is provided free of charge but you are required to read our [disclaimer](#) and to [cite](#) us when publishing results.

Download: [Network 10](#) [Network 4](#) [release notes](#) [user guide](#)

Recommended add-ons: [DNA Alignment](#) [Network Publisher](#)

[Go to SARS-CoV-2](#) early outbreak phylogeographic network page (PNAS April 2020)

Network 10 is 2MB in size and runs on Windows 10, 8, and 7. The user guide (pdf) is 2MB in size. Download the Network zip file, unzip to a folder of your choice, and double-click Network.exe to run. Network can run from memory stick - no Windows registry entries and no Windows Administrator required for installation.

Network Publisher is recommended for publication quality graphics, more pie chart network functions, display of longer node names (15 characters) and link names (9 characters), network root proxy nodes, mutations per link and individuals per node, and for diagram comparison.

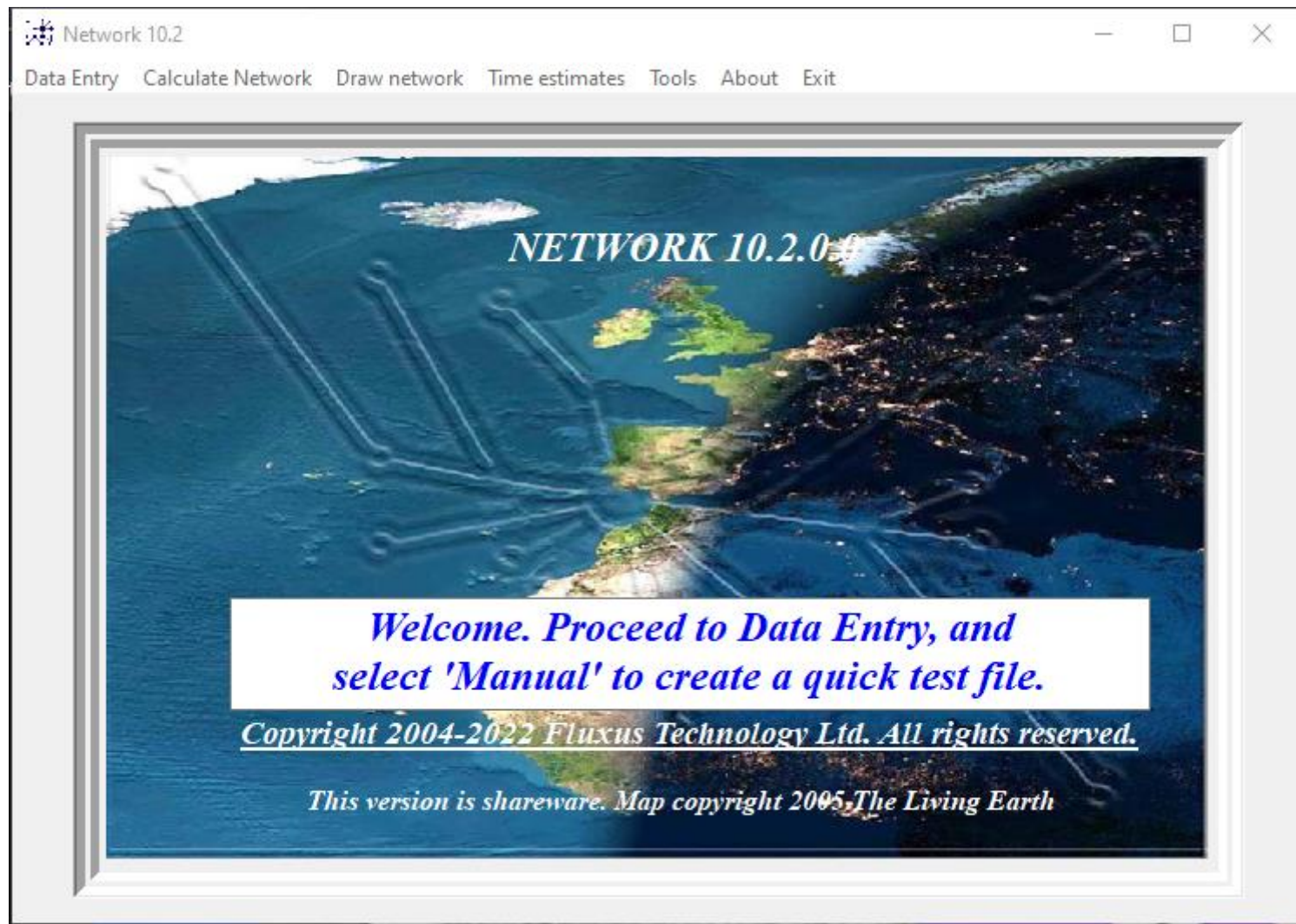
DNA Alignment is recommended for importing Sanger sequences (dna) or amino acid sequences into Network, helping you [avoid errors](#).

Network © Copyright Fluxus Technology Ltd 1999-2024. Web page last updated on 19 December 2023

Keywords: median joining, reduced median, forward simulation, outgroup rooting, [Bandelt](#), [Forster](#), [Röhl](#)

The Network software was developed to reconstruct *all possible shortest least complex* phylogenetic trees (all *maximum parsimony* or *MP* trees) from a given data set. Two different network-building options are included which can be used independently of each other.

Network



Statistical Parsimony

Statistical parsimony network

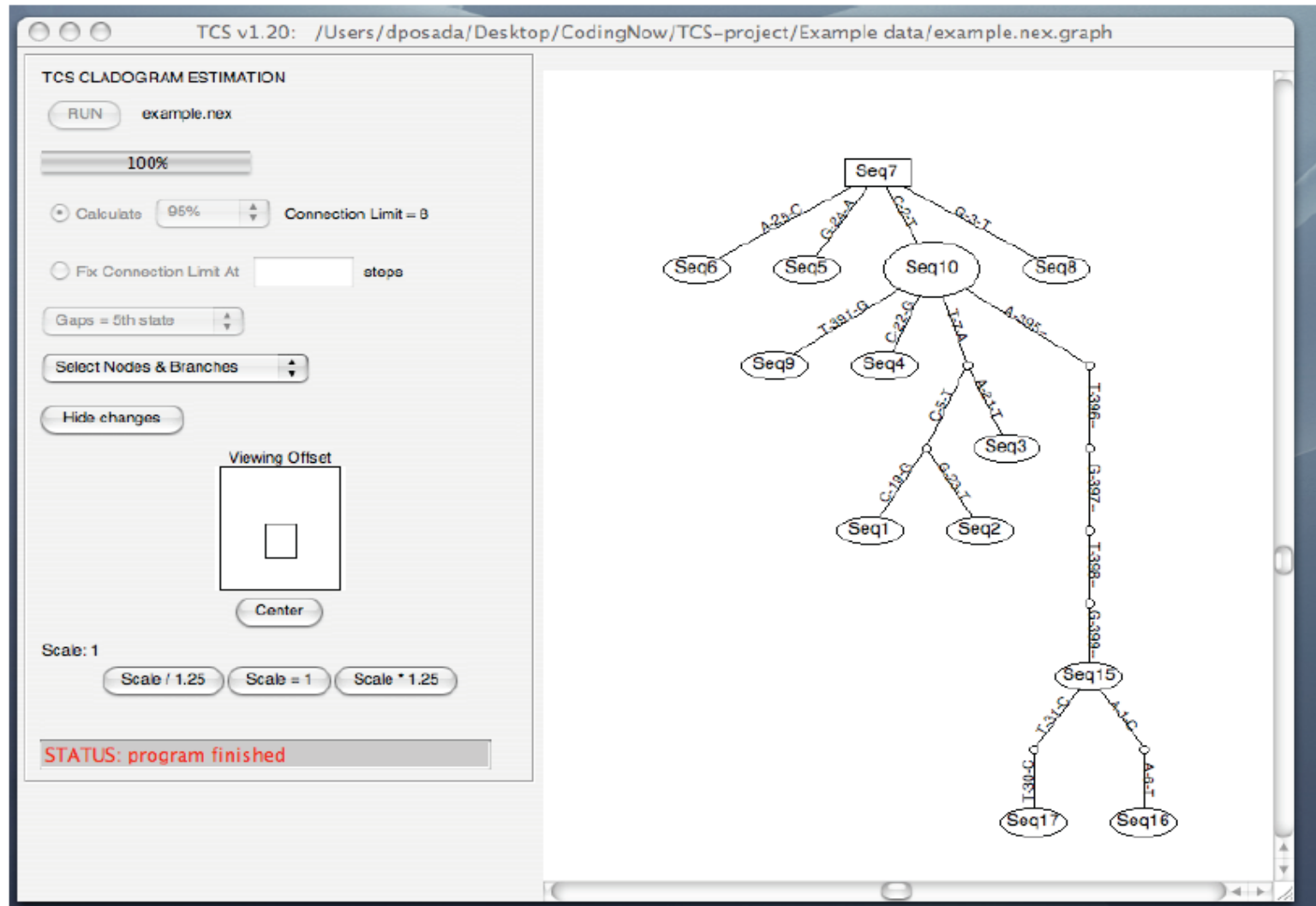
The method (Templeton et al., 1992; implemented in TCS, v. 1.13 [Clement et al., 2000]) first defines the uncorrected distance above which the parsimony principle is violated with more than 5% probability ("parsimony limit"). Then, all connections are iteratively established among haplotypes starting with the smallest distances and ending either when all haplotypes are connected or the distance corresponding to the parsimony limit has been reached. Although missing node haplotypes can be inferred using the TCS program, the exact algorithm is not described in the literature yet. For graph construction using the TCS program.

Statistical Parsimony

The statistical parsimony algorithm begins by estimating the maximum number of differences among haplotypes as a result of single substitutions (i.e. those that are not the result of multiple substitutions at a single site) with a 95% statistical confidence.

This number is called the parsimony limit (or parsimony connection limit). After this, haplotypes differing by one change are connected, then those differing by two, by three and so on, until all the haplotypes are included in a single network or the parsimony connection limit is reached. The statistical parsimony method emphasizes what is shared among haplotypes that differ minimally rather than the differences among the haplotypes and provides an empirical assessment of deviations from parsimony.

TCS



PopART <https://popart.maths.otago.ac.nz/>

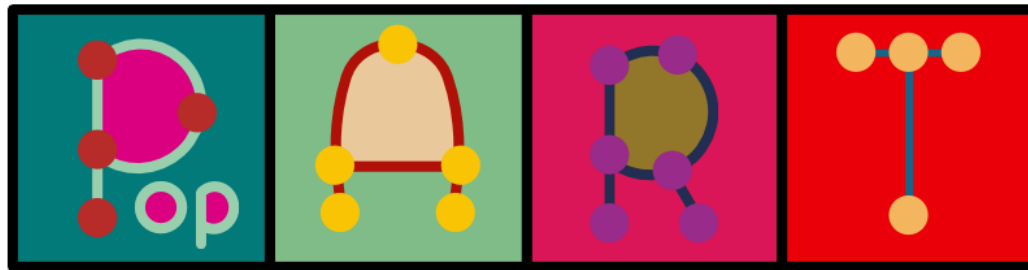


University of Otago Popart

[Home](#) [Download](#) [How To Cite](#) [Documentation](#) [Screenshots](#) [Changes](#) [Contact](#) [Links](#)



Search

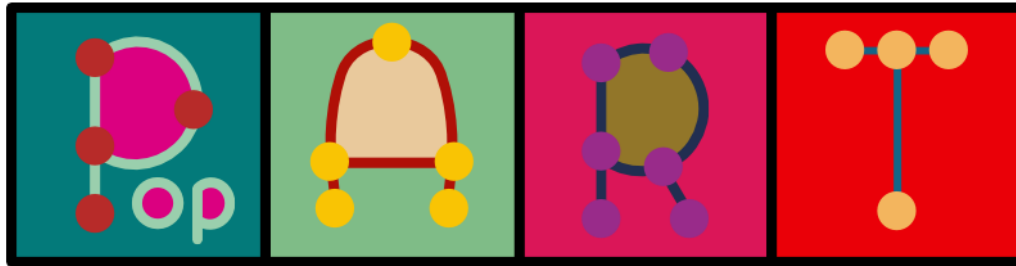


PopART (Population Analysis with Reticulate Trees) is free population genetics software that was developed as part of the [Allan Wilson Centre](#) Imaging Evolution Initiative. This is a collaborative project involving mathematicians and biologists from five universities and research institutes across New Zealand to develop better software to understand evolutionary relationships among populations.

The primary developer for PopART is [Jessica Leigh](#). Other contributors are [David Bryant](#) and [Mike Steel](#), as well as a dedicated group of beta-testers.

Leigh & Bryant 2015 MEE

PopART



By downloading PopART, you accept the [terms and conditions](#) for using PopART. PopART is distributed under the [Lesser GNU Public License](#).

Please select the download appropriate for your operating system (OS)

Stable release downloads:

[PopART version 1.7 for Windows](#)

DOWNLOAD

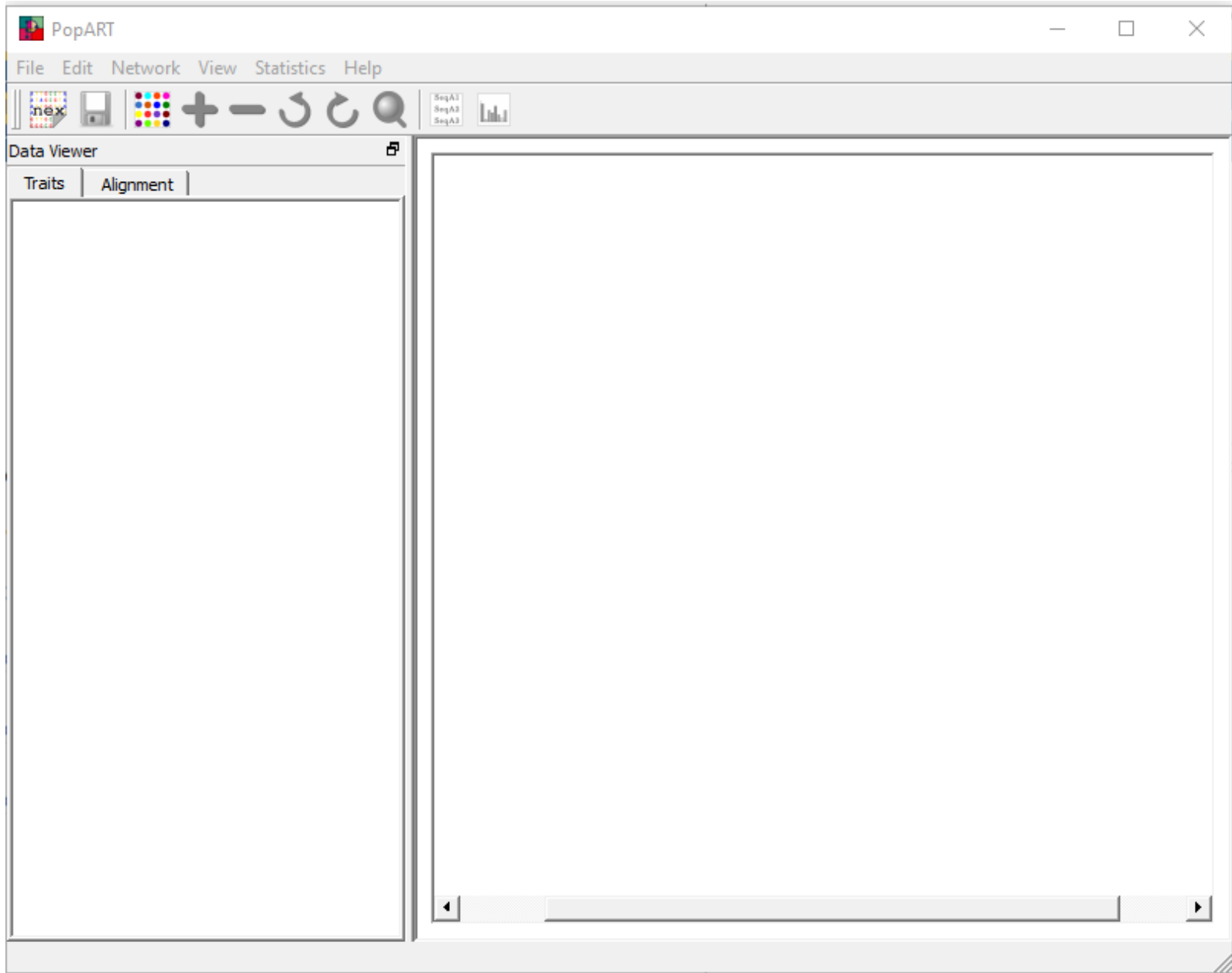
[PopART version 1.7 for Mac OSX \(≥ 10.6\)](#)

DOWNLOAD

[PopART version 1.7 sources for Linux/Unix/FreeBSD](#)

DOWNLOAD

PopART



PopART

The Networks

PopART includes several network inference methods. Minimum Spanning, Median Joining, and TCS are all popular methods that have been described elsewhere. Ancestral Parsimony, Integer Neighbor-Joining, and Tight Span Walker are new methods.

PopART

Ancestral Parsimony

This method shows different ways the haplotype sequences might be connected through ancestral sequences inferred by the Maximum Parsimony criterion. You'll find that these networks tend to have a lot of edges, because there's often a lot of uncertainty in the inference of ancestral sequences. You can reduce this somewhat by increasing the "Minimum Ancestral Frequency" value.

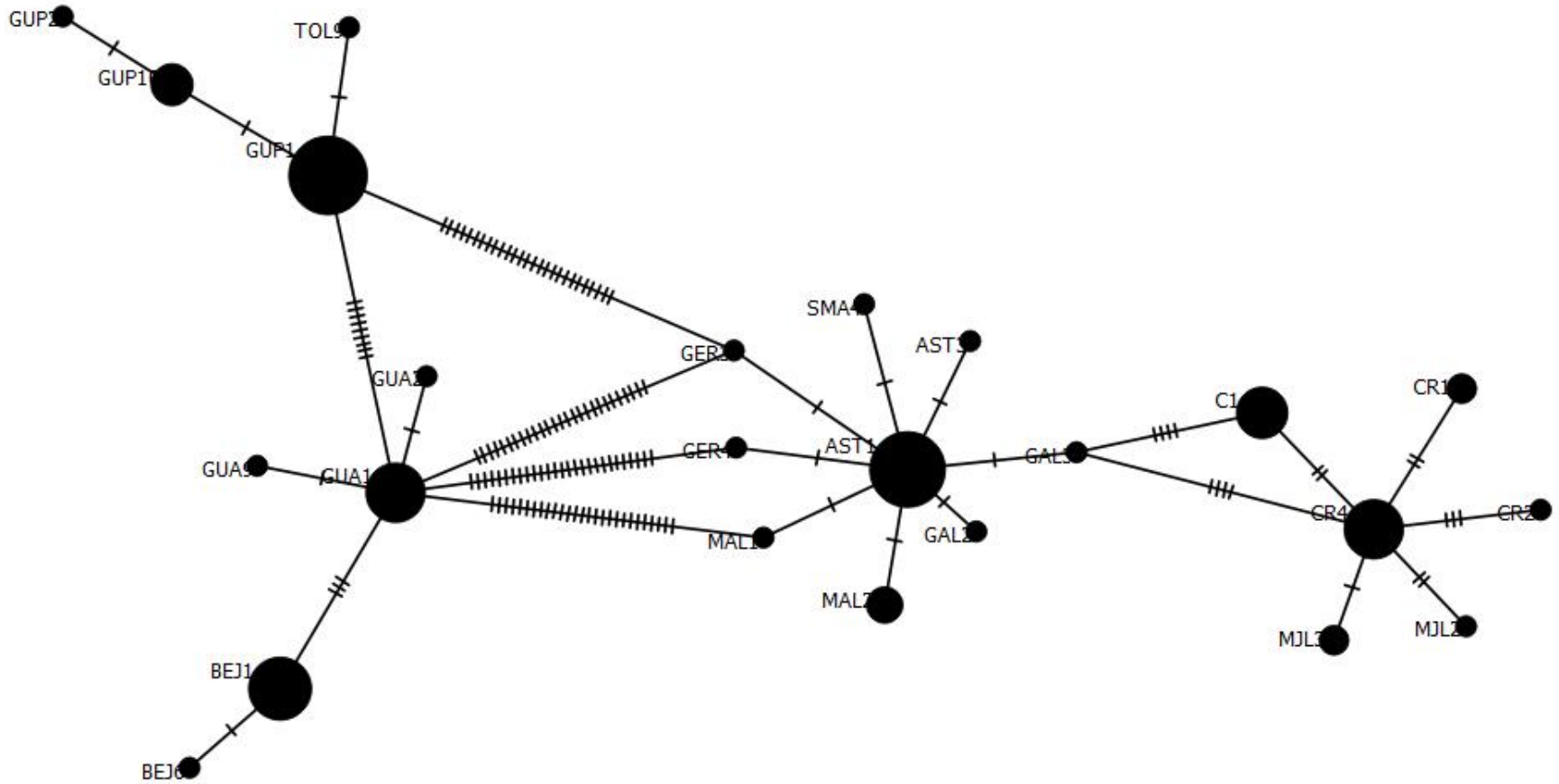
Integer Neighbor-Joining

This method begins with a Neighbor-Joining tree, but sets the branch lengths to integer values so that they represent the number of mutations between sequences. Then, for pairs of haplotypes whose distances on the tree are longer than the distances between the sequences, edges are added to shorten the distance. New edges won't be added if they create any distances on the network that are shorter than the distances between sequences. If a new edge is longer than the improvement produced by its addition (i.e., the difference between the distance on the network and the sequence distance), it won't be added. You can increase the "Reticulation tolerance" to allow the addition of these edges if they aren't much too long.

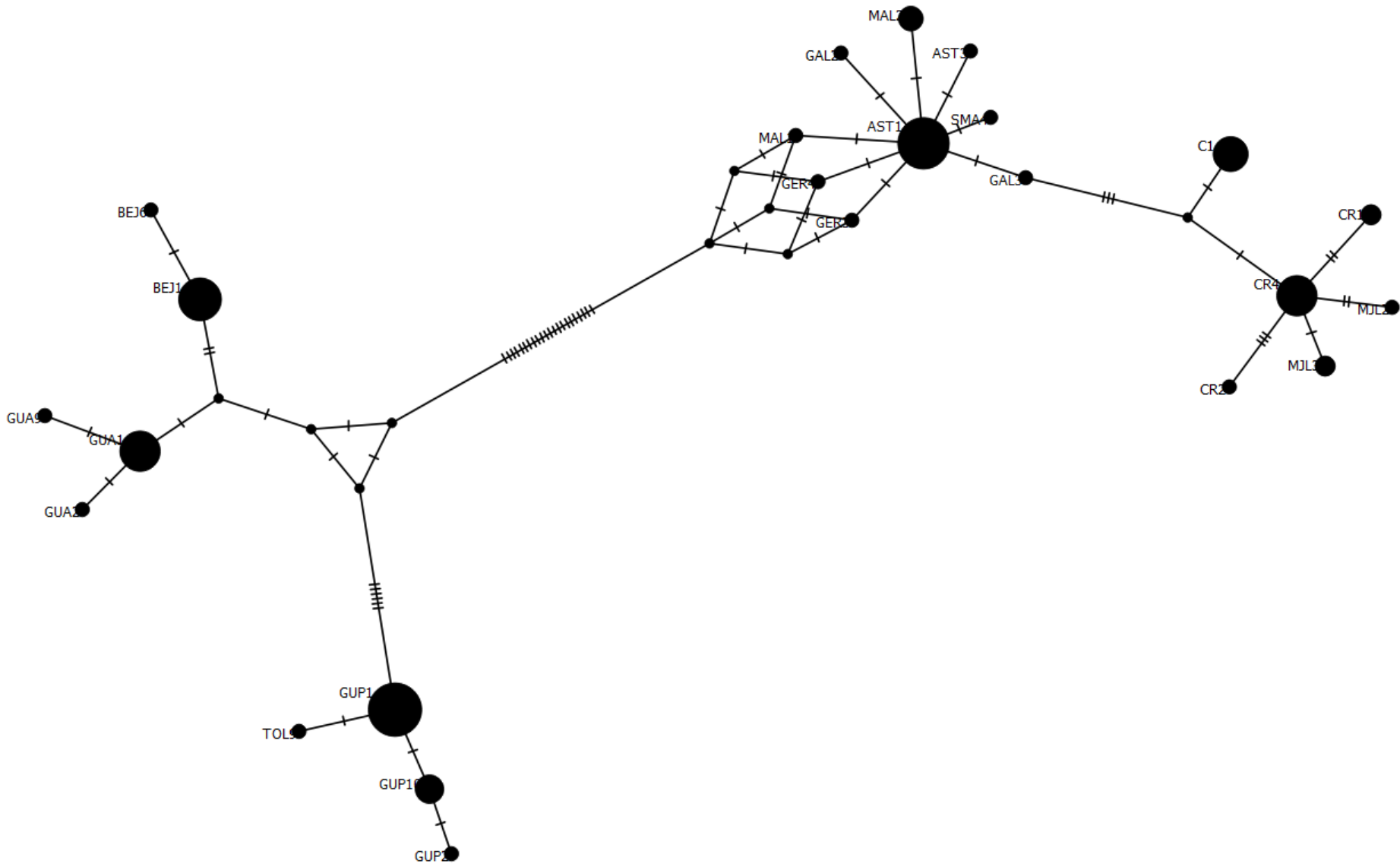
Tight Span Walker

This method uses some properties of the tight span for the distance matrix for your sequences to build a haplotype network. We'll be describing this method in detail soon, but it's beyond the scope of this documentation. The current implementation will fail with some data sets, but if it works, the distances on the network should match your sequence distances exactly.

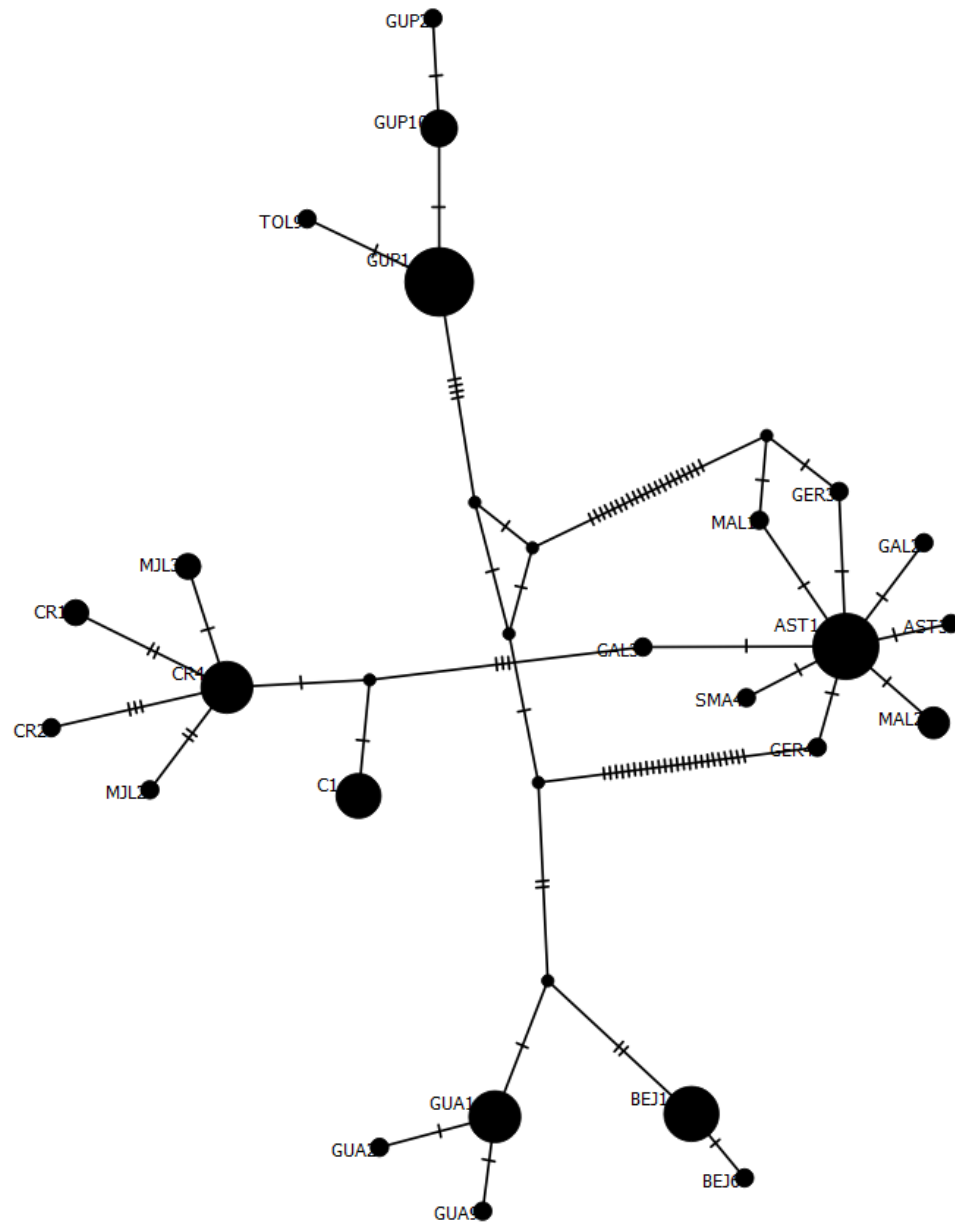
PopART - MSN



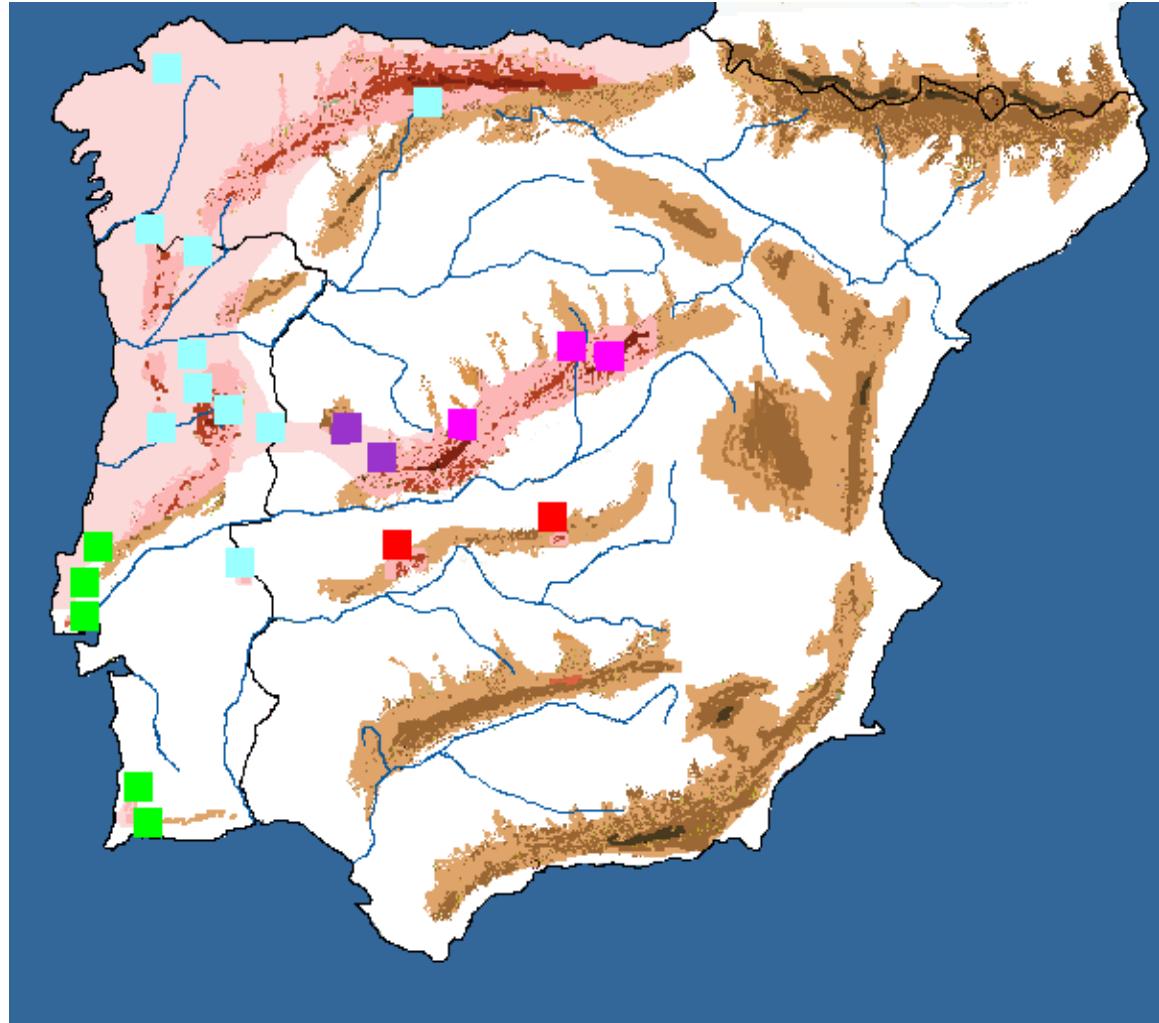
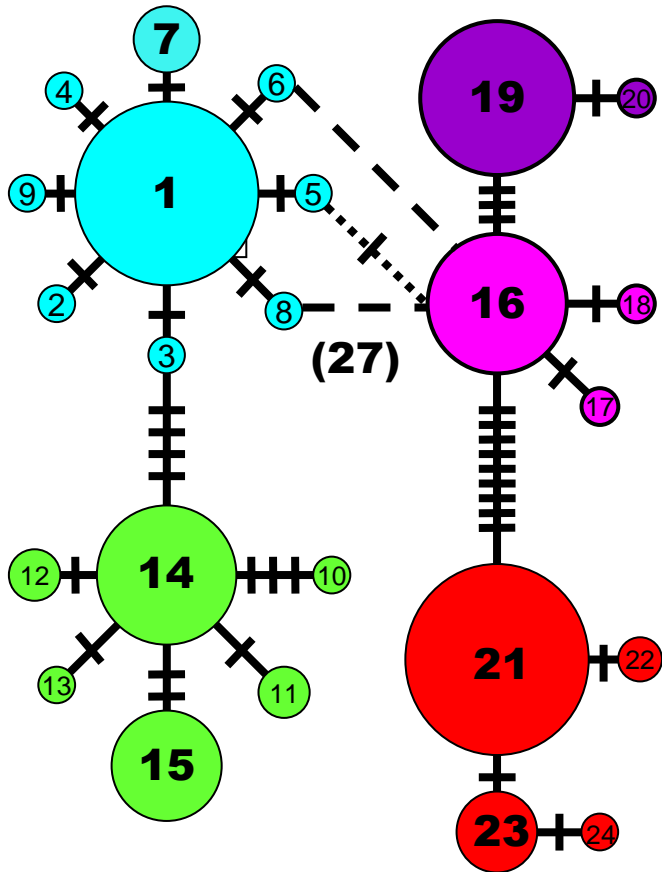
PopART Median Joining Network



PopART - TCS



Minimum spanning network of the 24 haplotypes found in the populations of *L. schreiberi*



Received: 4 November 2022 | Revised: 4 June 2023 | Accepted: 19 June 2023

DOI: 10.1111/1755-0998.13829

RESOURCE ARTICLE

MOLECULAR ECOLOGY
RESOURCES WILEY

fastHaN: a fast and scalable program for constructing haplotype network for large-sample sequences

Lianjiang Chi^{1,2} | Xiaolong Zhang^{1,2,3} | Yongbiao Xue^{1,2,3}  | Hua Chen^{1,2,3,4} 

FastHaN

Received

DOI: 10.

RESO

fast
hap

Lianji

Abstract

Haplotype networks can be used to demonstrate the genealogical relationships of DNA sequences within species, and thus are widely applied in population genetics, molecular ecology, epidemiology and evolutionary studies. However, existing programs become computationally infeasible as the sample size increases. Here, we present fastHaN, an efficient and scalable program suitable for constructing haplotype networks for large samples. On a data set with the haplotype length of 30 kb, the Median Joining Network (MJN) algorithm implemented by fastHaN is 3000 times faster than PopART and 70 times faster than NETWORK in single-threaded mode. The implementation of the Templeton-Crandall-Sing (TCS) algorithm is 100 times faster than PopART and 5800 times faster than the TCS software. Moreover, fastHaN also enables multi-threaded mode with scalability. The source code is freely available on <https://github.com/ChenHuaLab/fastHaN/>. A web-based version is also available on <https://ngdc.cncb.ac.cn/haplotype/>.

KEYWORDS

computational efficiency, gene genealogy, haplotype network, MJN algorithm, multi-threaded, TCS algorithm



Received
DOI: 10.
RESO

fast
hap
Lianji

ILEY

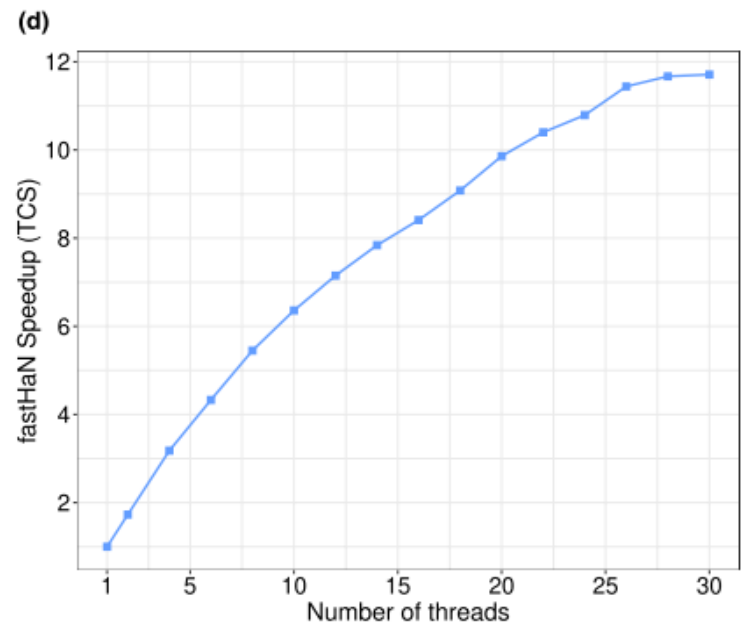
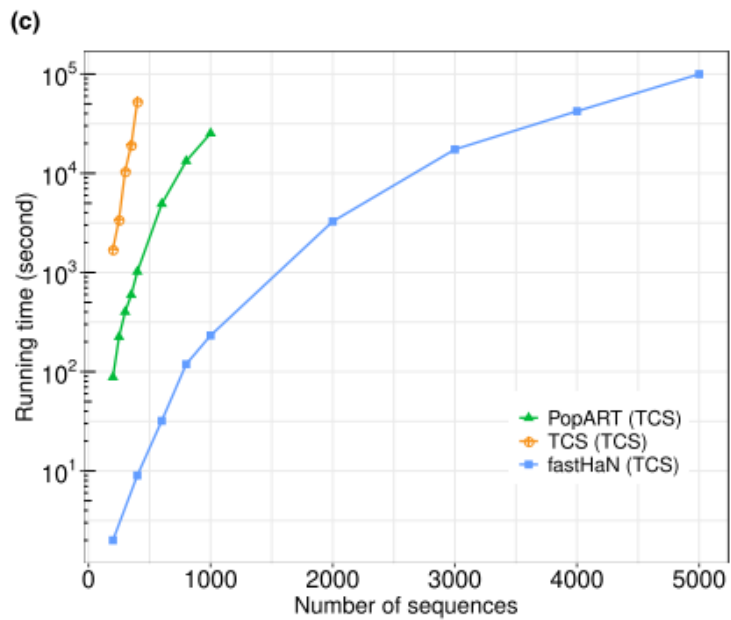
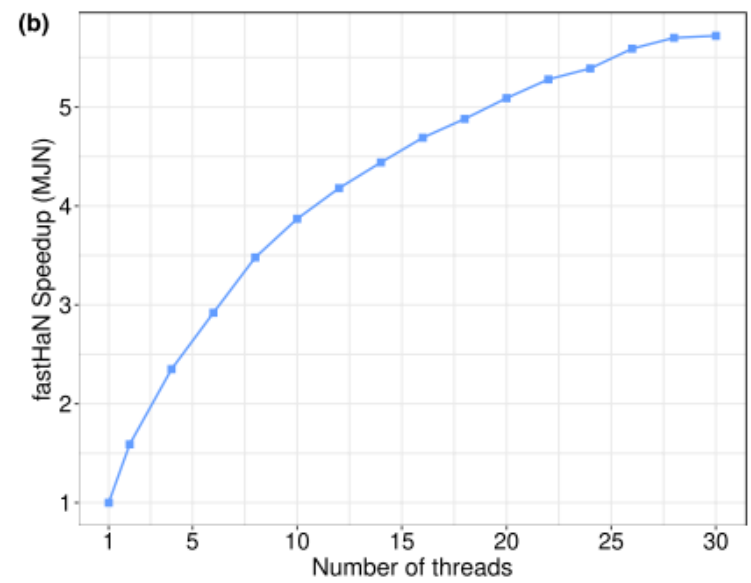
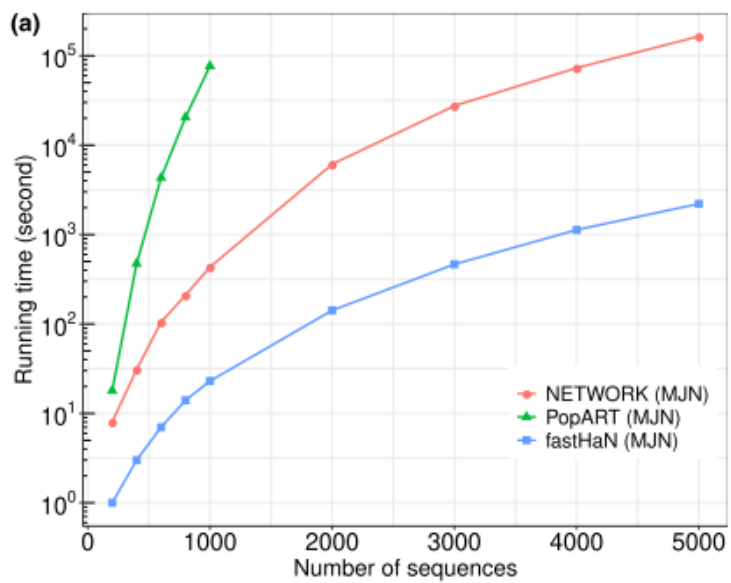


FIGURE 1 (a) Comparison of running time for MJN algorithm implemented by PopART (v.1.7), Network (v.10.2.0.0) and fastHaN on data with different sample sizes. (b) The speedup plot of fastHaN for the paralleled MJN algorithm. (c) Comparison of running time for TCS algorithm implemented by TCS (v.1.21), PopART (v.1.7) and fastHaN on data with different sample sizes. (d) The speedup plot of fastHaN for the paralleled TCS algorithm.