

HaploThread: a scalable desktop tool for efficient haplotype network inference and interactive visualization

Bo Xu ^{1,2†}, Lun Li ^{1,2,3†}, Cuiping Li ^{1,2,3,4}, Anke Wang ^{1,2,3,4},
Zhuojing Fan ^{1,2}, Shuhui Song ^{1,2,3,5*}

¹National Genomics Data Center, China National Center for Bioinformation, Beijing 100101, China

²Beijing Institute of Genomics, Chinese Academy of Sciences, Beijing 100101, China

³Beijing Key Laboratory of Intelligent Governance and Application of Biological Big Data, China National Center for Bioinformation, Beijing 100101, China

⁴University of Chinese Academy of Sciences, Beijing 100049, China

⁵Sino-Danish College, University of Chinese Academy of Sciences, Beijing 100049, China

*Corresponding author: E-mail: songshh@big.ac.cn.

[†]These authors contributed equally.

Associate editor: Haipeng Li

Abstract

This note introduces HaploThread, a user-friendly desktop software with a graphical user interface designed for constructing and visualizing haplotype networks. Developed in C++ using the Qt library, HaploThread integrates network visualization with multiple multithreaded haplotype construction algorithms—including McAN and fastHaN (which incorporates MSN, MJN, and TCS)—through a modular plugin architecture. It provides an intuitive workflow for building and visualizing haplotype networks from large-scale datasets, while also supporting functional extensions via plugins to facilitate the analysis of genetic variation and evolutionary relationships. HaploThread is released as open-source software under the GNU General Public License. Its source code and precompiled executables for Windows and macOS are freely available at <https://ngdc.cncb.ac.cn/biocode/tool/BT007948> and <https://github.com/git-xubo/HaploThread>.

Keywords visualization, network construction, multithreading, software

Introduction

Desktop software plays a crucial role in biological research by providing powerful computational tools with user-friendly graphical interfaces, enabling researchers to analyze complex biological data efficiently (Kumar and Dudley 2007). Favored by biologists, it offers a balanced solution that combines ease of use, high performance, and data security—unlike command-line tools that require programming skills or web-based platforms that depend on internet access.

Desktop software for haplotype network analysis is particularly significant in population genetics for inferring evolutionary relationships, tracking disease transmission, and studying population dynamics (Song et al. 2020; Li et al. 2024). However, traditional state-of-art desktop software for haplotype network construction and/or visualization, including TCS (Clement et al. 2000), Network (<https://www.fluxus-engineering.com/>), PopART (Leigh and Bryant 2015), HapStar (Teacher and Griffiths 2011) and Hapsolutely (Vences et al. 2024), only integrated single-threaded algorithms (Templeton et al. 1992; Excoffier and

Smouse 1994; Bandelt et al. 1999; Matschiner 2016), which limits their efficiency when processing large datasets (Table 1).

Currently known multithreaded construction algorithms include McAN (Li et al. 2023) and fastHaN (Chi et al. 2023), which are faster than all existing algorithms even in single-threaded mode. Among them, McAN is currently the only program capable of processing sequences on the order of millions. Nevertheless, McAN provides only a simple command-line interface, which may appear daunting to beginning users, and an online web service, which depends on internet connection availability and stability and may bring concerns about privacy and security risks of user data.

To address these challenges, we developed HaploThread, a fully local software that integrates intuitive user interfaces with advanced multithreaded algorithms. This ensures ease of use, data security, and computational efficiency in haplotype network construction and visualization. Designed with practical research workflows in mind, HaploThread is primarily intended for wet-lab researchers, offering an intuitive interface for rapid network construction and visualization. At the same time, it also serves

Received: August 21, 2025. **Revised:** January 12, 2026. **Accepted:** February 9, 2026

© The Author(s) 2026. Published by Oxford University Press on behalf of Society for Molecular Biology and Evolution.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (<https://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact reprints@oup.com for reprints and translation rights for reprints. All other permissions can be obtained through our RightsLink service via the Permissions link on the article page on our site—for further information please contact journals.permissions@oup.com.

Table 1 Comparison of programs for haplotype network construction and/or visualization.

Software	Input formats	Output formats	Multithreading	Scalability	Algorithms	Language	Layout	Temporal dynamic view	Map view	Other features	URL
HaploThread	VCF, Phylip, GML, GraphML	GML, GraphML, PDF, SVG, PNG, JPEG, BMP	Yes	Yes	McAN, MSN, MJN, TCS	C++, JS	Adjust layout in real time	Yes	Yes	...	https://ngdc.cncb.ac.cn/biocode/tool/BT007948
HapNetworkView	VCF, GML, Phylip	SVG, PDF, PNG	Yes	No	MJN	Java	Adjust layout in real time	No	No	...	https://github.com/HapNetworkView/HapNetworkView
PopART	Nexus, Phylip	PDF, SVG, PNG	No	No	AMP, INJ, MSN, MJN, TCS, TSW	C++	Adjust only one node manually at a time	No	Yes	...	https://popart.maths.otago.ac.nz/
Hapsolutely	Fasta, TSV	PDF, SVG, PNG	No	No	MSN, MJN, TCS, TSW, FTN	Python	Adjust only one node manually at a time	No	No	Phasing	https://github.com/iTaxoTools/Hapsolutely
Network	RDF, YCH, SCO, Phylip, Nexus, Fasta/VCF ^a	PDF, BMP	No	No	MJN, RM	Unknown ^b	Adjust only one node manually at a time	No	No	...	https://www.fluxus-engineering.com/sharenet.htm
TCS	Phylip, Nexus	None ^c	No	No	TCS	Java	Adjust only one node manually at a time	No	No	...	https://bioresearch.byu.edu/tcs/
HapStar	RTF, TXT	SVG	No	No	None	Python	Manually click a button to adjust the layout	No	No	...	https://fo.am/activities/hapstar/
McAN Web Platform	VCF, TSV	PNG	Yes	No	McAN	C, C++, JS	Adjust layout in real time	Yes	Yes	...	https://ngdc.cncb.ac.cn/bit/hapnet/
fastHan	Phylip	None ^d	Yes	No	MSN, MJN, TCS	C	None	No	No	...	https://ngdc.cncb.ac.cn/haplotype/

^aProvide tools for converting Fasta/VCF to RDF.
^bThis software is not open source.
^cImages cannot be exported directly but can be displayed in software.
^dThe GML files it generates need to be visualized with tcsBU.

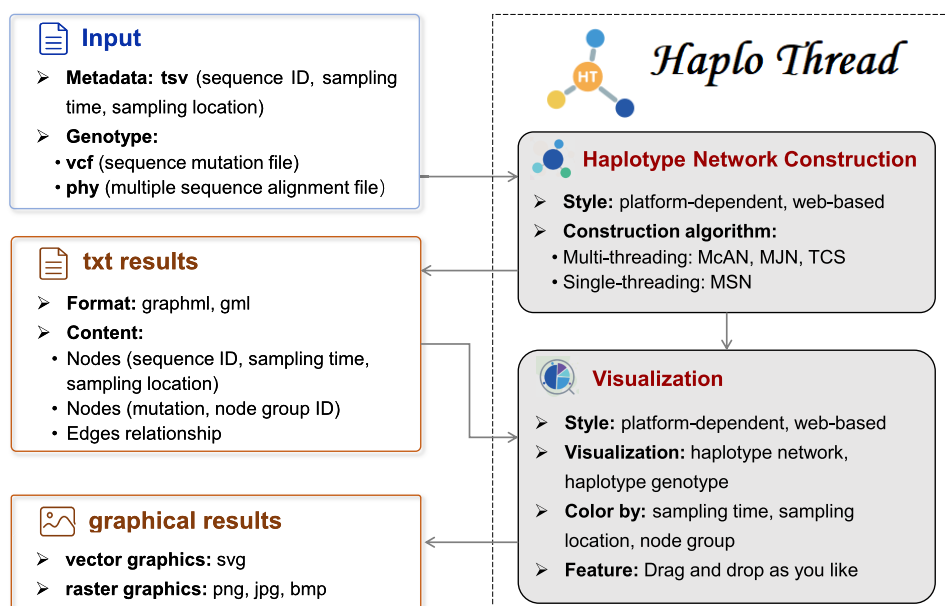


Figure 1 Workflow of HaploThread, illustrating the network construction module and the network visualization module, together with their major functionalities, input data, and output representations.

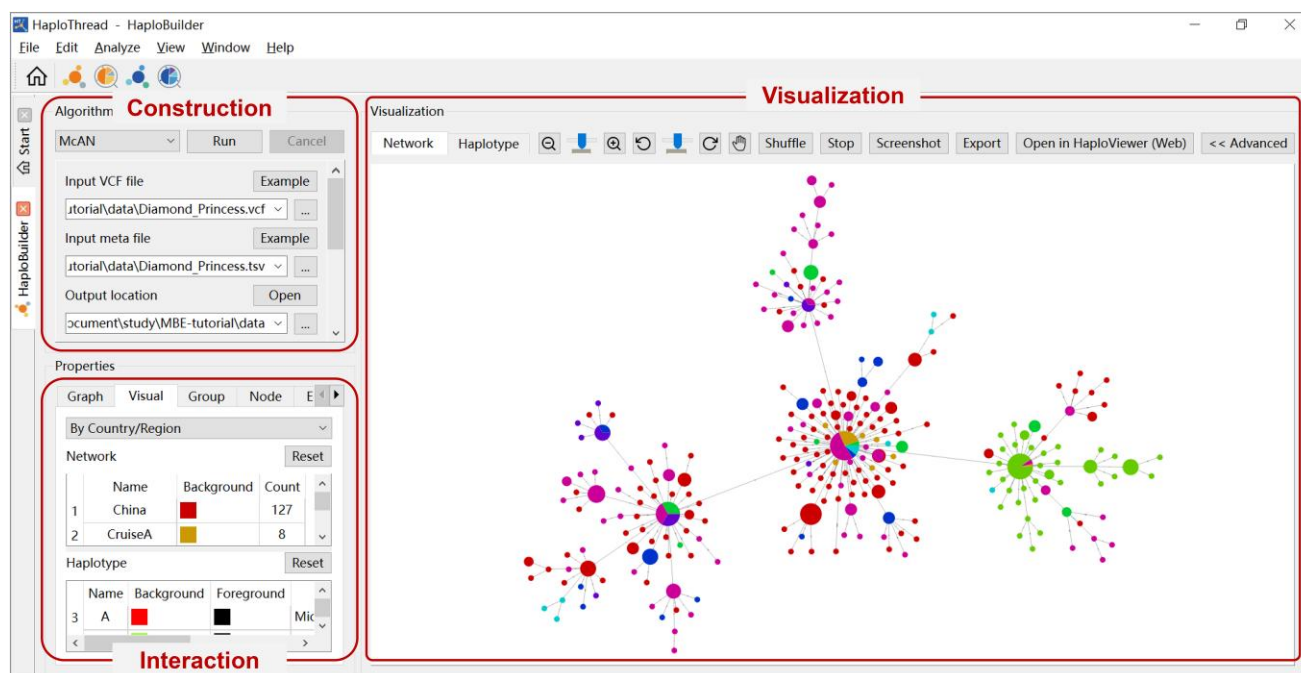


Figure 2 Screenshot of HaploThread showing the haplotype construction and visualization modules in the platform-dependent interface style. Nodes are colored based on group assignments.

bioinformaticians by providing fast algorithms and interactive visualizations that streamline exploratory analyses and enable efficient interpretation of large and complex haplotype networks. Built upon McAN, an algorithm from our previous work, and enhanced with the advanced features of fastHaN, HaploThread leverages the performance advantages of desktop environments and multithreading to deliver rapid, scalable, and reliable analyses for evolutionary studies in population genetics.

Functionality

HaploThread currently comprises two functional modules: the network construction module and the network visualization module (Fig. 1; Figure S1). Each module is available in two distinct visual style of interface: a platform-dependent style (Fig. 2; Figure S2) and a web-based style (Figures S3 and S4).

The construction module processes sequence data in either VCF or PHYLIP format and generates the constructed network

file in GraphML or GML format. Both styles of interfaces offer identical functionality, ensuring a consistent user experience. Users have the flexibility to choose a multithreaded construction algorithm from McAN, TCS, MSN, or MJN, specify the required parameters, configure the number of computational threads based on available system resources, and start the process by clicking the “Run” button. The entire workflow is designed to be intuitive and user-friendly, requiring minimal explanation for efficient operation.

The visualization module is designed to open network files in GraphML or GML formats and employs a force-directed algorithm to layout the network. If users provide metadata, such as sampling date, geographic location, or clustering information, via a metadata file, all samples and network nodes are automatically assigned distinct colors according to the associated metadata, generating visualization color schemes that are available for user selection and application (Figures S5 and S6). Additionally, users can observe the temporal dynamics of the network by either automatically running or manually dragging the timeline (Figure S7). The constructed network can be converted between GraphML and GML formats and exported as a PDF file or images in various formats, such as SVG, PNG, JPEG, or BMP.

The functionalities of the two styles of visualization differ slightly. The platform-dependent interface provides a visual browser for haplotype sequence differences, which can be displayed and selected in synchronization with the haplotype network. Users can customize the color and font attributes of visual elements to suit their preferences. In contrast, the web-like interface offers a map browsing function, allowing nodes to be displayed on a world map when country-level metadata is available.

Implementation

HaploThread software is written using the Qt cross-platform application framework (Figure S8). Compiled executables for Windows and macOS are provided. It is architected with a plugin-based approach. Functional modules are designed in the form of plugins and can be integrated, updated, or uninstalled independently of the core application, reflecting the flexibility and extensibility of the software.

The haplotype network construction module in HaploThread is implemented by integrating the software McAN and fastHaN into a plugin. The platform-dependent style of the module is developed using C++ and built on the Qt Widgets module. This approach ensures high performance and enhanced flexibility for desktop applications and seamless consistency with the native styles of operating systems such as Windows and macOS. The web-based style of the module is written using JavaScript and HTML, and it is embedded within the plugin via the Qt WebEngine module. This design provides users with an experience that is consistent with online web services.

Future improvements may involve the implementation of novel algorithms for network construction and layout, 3D visualization of networks, and expanding support for additional input and output formats.

Performance comparison

Speed comparison

To evaluate the runtime performance of different haplotype network construction tools, we conducted a benchmarking

experiment using SARS-CoV-2 sequences downloaded from the RCoV19 database. Three datasets were prepared, containing 500, 1,000, and 5,000 sequences, respectively (Data S1). We tested five representative state-of-the-art desktop software tools: PopART, HaploThread, Hapsolutely, Network, and HapNetworkView. For a fair comparison, all tools were run in single-threaded mode regardless of their support for parallel computing.

To reflect typical usage scenarios for desktop applications, all tests were conducted on a consumer-grade laptop running Windows 10 Home Edition, equipped with an Intel Core i7-10510U CPU (1.80 GHz, four cores, eight logical processors), 16 GB RAM, and a 512 GB SSD.

The results demonstrate that HaploThread consistently outperformed all other tools in terms of execution time across all three datasets (Table S1). For tools that offer multiple algorithms, only the fastest configuration was included in the comparison. Notably, on the 5,000-sequence dataset, HaploThread completed both network construction and visualization in just 23 s, while PopART, Hapsolutely, and Network were unable to finish the task within 1 h. This highlights the superior efficiency and scalability of HaploThread in handling large-scale haplotype data under real-world computing conditions.

Accuracy evaluation

To further assess the reliability of HaploThread, we evaluated its performance using the *Diamond Princess* dataset (see Data S2), which was obtained from Sekizuka et al. (2020). The widely used PopART software (implementing the MJN, TCS, and MSN algorithms among other algorithms) and the command-line version of McAN were adopted as benchmark tools for comparison.

The spectral feature similarity (see Note S1) was employed as the evaluation metric. This measure has been broadly applied in network comparison tasks due to its high computational efficiency, sensitivity to topological structure, and independence from node labeling, making it particularly suitable for evaluating the structural consistency of haplotype networks.

We computed the spectral feature similarity between haplotype networks reconstructed by the four algorithms implemented in HaploThread and those generated by the corresponding algorithms in PopART and McAN. The results showed that all four HaploThread-generated networks exhibited a similarity score of 1.00 with their respective benchmark networks (similarity range: 0 to 1), demonstrating excellent consistency and algorithmic reliability.

To further confirm statistical significance, we compared the similarities between randomly generated networks and benchmark networks. HaploThread-generated networks exhibited significantly higher similarity than random networks ($P < 0.001$; see Figure S9), with detailed procedures provided in Note S2.

Together, these results demonstrate the reliability of HaploThread for accurately reconstructing haplotype networks.

Case study: Diamond Princess SARS-CoV-2 outbreak

To illustrate the application of haplotype network analysis for investigating viral origins, we analyzed SARS-CoV-2 genomes from the Diamond Princess (DP) outbreak using HaploThread.

Previous studies have predominantly supported a single introduction scenario, based on the observation that all DP-associated haplotypes share the defining mutation G11083T (Sekizuka et al. 2020). At the same time, other studies have reported more complex transmission dynamics during shipboard quarantine, raising the possibility of alternative evolutionary pathways (Yeh and Contreras 2021).

We reconstructed haplotype networks using four algorithms implemented in HaploThread (McAN, MSN, MJN, and TCS). Notably, all four methods consistently recovered the evolutionary path defined by the G11083T substitution linking the DP cluster to the Wuhan-Hu-1 reference, supporting the predominant single-introduction hypothesis reported previously (Figures S10 to S13). The McAN network, which outputs a minimum-cost arborescence (or a rooted directed tree), retained only this most parsimonious connection. In contrast, networks inferred using MSN, MJN, and TCS preserved additional alternative paths beyond G11083T, revealing increased topological complexity. These additional connections do not contradict the shared origin defined by G11083T but instead highlight potential recurrent substitutions, recombination, or unsampled intermediates, as suggested by studies reporting within-outbreak diversification.

Together, these results demonstrate that haplotype networks represent one line of evidence for viral origin inference and that comparative analysis across reconstruction algorithms helps distinguish robust evolutionary signals from method-dependent network features.

Discussions and conclusion

The rapid emergence and evolution of novel and outbreak-associated pathogens pose significant challenges to public health surveillance and epidemic control. Timely and accurate construction of haplotype networks is essential for tracking pathogen evolution, understanding transmission dynamics, and informing intervention strategies. Traditional haplotype network construction GUI desktop software often struggles to handle large-scale genomic datasets efficiently, limiting their applicability in urgent outbreak scenarios. HaploThread was specifically designed to address that problem.

Currently, a state-of-the-art desktop software for haplotype network construction and visualization, named HapNetworkView (Chi et al. 2025), has been developed. It integrates a multithreaded MJN algorithm from fastHaN. In comparison, HaploThread offers broader advantages in terms of algorithmic diversity. In addition to supporting a multithreaded MJN algorithm, HaploThread integrates five state-of-the-art multithreaded haplotype network construction algorithms, including McAN, MJN, MSN, and two implementations of TCS. This comprehensive algorithm suite provides users with enhanced flexibility and computational efficiency, especially when analyzing large-scale datasets, making HaploThread a versatile and powerful tool for haplotype network analysis. In addition to its algorithmic strengths, HaploThread also offers advanced visualization capabilities that enhance interpretability and user interaction. HaploThread not only enables coloring of nodes in the haplotype network based on geographic information but also supports the dynamic display or concealment of nodes according to sampling time. These spatiotemporal visualization capabilities provide users with enhanced flexibility to explore evolutionary relationships and track the temporal and

geographic spread of haplotypes, which is especially valuable in infectious disease research.

In summary, HaploThread simplifies the complex process of haplotype network analysis by integrating multiple methods of construction and visualization into a unified, user-friendly platform. Its intuitive design ensures accessibility for users with varying levels of expertise. Furthermore, the tool's ability to handle large datasets, combined with its support for plugin mechanisms, facilitates community-driven improvements and customization, enhancing its adaptability to evolving research needs.

Supplementary material

Supplementary material is available at *Molecular Biology and Evolution* online.

Funding

This work was supported by the Key Collaborative Research Program of the Alliance of National and International Science Organizations for the Belt and Road Regions (grant no. ANSO-CR-KP-2022-09), the Biological Breeding-National Science and Technology Major Project (grant no. 2022ZD04017), and the National Natural Science Foundation of China (grant nos. 32270718, 32170678).

Conflicts of interest

The authors declare that they have no conflict of interest.

Data availability

The datasets used for performance benchmarking in this study—comprising subsets of 500, 1,000, and 5,000 SARS-CoV-2 sequences from the RCoV19 database—are provided in Data S1. The dataset used in Fig. 2 is provided in Data S2. The tutorial of HaploThread is provided in Data S3.

Code availability

Source code and installation packages of HaploThread are freely available at <https://ngdc.cncb.ac.cn/biocode/tool/BT007948> and <https://github.com/git-xubo/HaploThread> under the GNU General Public License (GPL).

References

- Bandelt HJ, Forster P, Röhl A. Median-joining networks for inferring intraspecific phylogenies. *Mol Biol Evol.* 1999;16:37–48. <https://doi.org/10.1093/oxfordjournals.molbev.a026036>.
- Chi L et al. HapNetworkView: a tool for haplotype network exploration and visualization. *BMC Genomics.* 2025;26:52. <https://doi.org/10.1186/s12864-025-11206-8>.
- Chi L, Zhang X, Xue Y, Chen H. Fasthan: a fast and scalable program for constructing haplotype network for large-sample sequences. *Mol Ecol Resour.* 2023;25:e13829. <https://doi.org/10.1111/1755-0998.13829>.

- Clement M, Posada D, Crandall KA. TCS: a computer program to estimate gene genealogies. *Mol Ecol*. 2000;9:1657–1659. <https://doi.org/10.1046/j.1365-294x.2000.01020.x>.
- Excoffier L, Smouse PE. Using allele frequencies and geographic subdivision to reconstruct gene trees within a species: molecular variance parsimony. *Genetics*. 1994;136:343–359. <https://doi.org/10.1093/genetics/136.1.343>.
- Kumar S, Dudley J. Bioinformatics software for biologists in the genomics era. *Bioinformatics*. 2007;23:1713–1717. <https://doi.org/10.1093/bioinformatics/btm239>.
- Leigh JW, Bryant D. Popart: full-feature software for haplotype network construction. *Methods Ecol Evol*. 2015;6:1110–1116. <https://doi.org/10.1111/2041-210X.12410>.
- Li L *et al*. McAN: a novel computational algorithm and platform for constructing and visualizing haplotype networks. *Brief Bioinform*. 2023;24:bbad174. <https://doi.org/10.1093/bib/bbad174>.
- Li L *et al*. Machine learning early detection of SARS-CoV-2 high-risk variants. *Adv Sci*. 2024;11:e2405058. <https://doi.org/10.1002/advs.202405058>.
- Matschiner M. Fitchi: haplotype genealogy graphs based on the Fitch algorithm. *Bioinformatics*. 2016;32:1250–1252. <https://doi.org/10.1093/bioinformatics/btv717>.
- Sekizuka T *et al*. Haplotype networks of SARS-CoV-2 infections in the Diamond Princess cruise ship outbreak. *Proc Natl Acad Sci U S A*. 2020;117:20198–20201. <https://doi.org/10.1073/pnas.2006824117>.
- Song S *et al*. The global landscape of SARS-CoV-2 genomes, variants, and haplotypes in 2019nCoV. *Genomics Proteomics Bioinformatics*. 2020;18:749–759. <https://doi.org/10.1016/j.gpb.2020.09.001>.
- Teacher AG, Griffiths DJ. HapStar: automated haplotype network layout and visualization. *Mol Ecol Resour*. 2011;11:151–153. <https://doi.org/10.1111/j.1755-0998.2010.02890.x>.
- Templeton AR, Crandall KA, Sing CF. A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping and DNA sequence data. III. Cladogram estimation. *Genetics*. 1992;132:619–633. <https://doi.org/10.1093/genetics/132.2.619>.
- Vences M *et al*. Hapsolutely: a user-friendly tool integrating haplotype phasing, network construction, and haploweb calculation. *Bioinform Adv*. 2024;4:vbae083. <https://doi.org/10.1093/bioadv/vbae083>.
- Yeh TY, Contreras GP. Viral transmission and evolution dynamics of SARS-CoV-2 in shipboard quarantine. *Bull World Health Organ*. 2021;99:486–495. <https://doi.org/10.2471/BLT.20.255752>.