Bioinformatics, 2025, **41(8)**, btaf393
https://doi.org/10.1093/bioinformatics/btaf393
Advance Access Publication Date: 12 July 2025
Applications Note

Genetics and population analysis

Tsbrowse: an interactive browser for ancestral recombination graphs

Savita Karthikeyan[1],*, Ben Jeffery[1], Duncan Mbuli-Robertson[1], Jerome Kelleher[1]

[1]Big Data Institute, Li Ka Shing Centre for Health Information and Discovery, University of Oxford, Oxford OX3 7LF, United Kingdom

*Corresponding author. Big Data Institute, Li Ka Shing Centre for Health Information and Discovery, University of Oxford, Old Road Campus, Oxford OX3 7LF, United Kingdom. E-mail: savita.karthikeyan@st-hughs.ox.ac.uk.

Associate Editor: Russell Schwartz

Abstract

Summary: Ancestral recombination graphs (ARGs) represent the interwoven paths of genetic ancestry of a set of recombining sequences. The ability to capture the evolutionary history of samples makes ARGs valuable in a wide range of applications in population and statistical genetics. ARG-based approaches are increasingly becoming a part of genetic data analysis pipelines due to breakthroughs enabling ARG inference at biobank-scale. However, there is a lack of visualization tools, which are crucial for validating inferences and generating hypotheses. We present `tsbrowse`, an open-source, web-based Python application for the interactive visualization of the fundamental building blocks of ARGs, i.e. nodes, edges and mutations. We demonstrate the application of `tsbrowse` to various data sources and scenarios, and highlight its key features of browsability along the genome, user interactivity, and scalability to very large sample sizes.

Availability and implementation: `Tsbrowse` is installed as a Python package from PyPI (https://pypi.org/project/tsbrowse/), while a development version is maintained at https://github.com/tskit-dev/tsbrowse. Documentation is available at https://tskit.dev/tsbrowse/docs/. Source code is archived on Zenodo with DOI, https://doi.org/10.5281/zenodo.15683039.

1 Introduction

Ancestral recombination graphs (ARGs) describe how a set of sample sequences relate to each other at each position along the genome in a recombining species, and are currently the subject of intense research interest (Brandt *et al.* 2024, Lewanski *et al.* 2024, Wong *et al.* 2024, Nielsen *et al.* 2025). ARGs are a fundamental object in population genetics, and although they have been of theoretical interest for decades (Hudson 1983, Griffiths and Marjoram 1996, 1997) it is only with recent breakthroughs in inference methods (Rasmussen *et al.* 2014, Kelleher *et al.* 2019, Speidel *et al.* 2019, Wohns *et al.* 2022, Zhang *et al.* 2023, Deng *et al.* 2024, Gunnarsson *et al.* 2024) that widespread application has become possible. Varied applications have been proposed, such as inferring selection (Stern *et al.* 2019, Hejase *et al.* 2022) and the spatial location of genetic ancestors (Deraje *et al.* 2024, Osmond and Coop 2024, Grundler *et al.* 2025), more powerful approaches to quantifying genetic relatedness (Fan *et al.* 2022, Zhang *et al.* 2023, Gunnarsson *et al.* 2024, Lehmann *et al.* 2025) and other methodological improvements for genome wide association studies (Link *et al.* 2023, Nowbandegani *et al.* 2023), and the development of machine learning methods using inferred ARGs as input (Hejase *et al.* 2022, Pearson and Durbin 2023, Korfmann *et al.* 2024, Whitehouse *et al.* 2024). While these developments are exciting, the performance of these new methods depends critically on the accuracy of the inferred ARGs. Although studies benchmarking the various inference methods on simulated data have emerged (Brandt *et al.* 2022, Deng *et al.* 2024, Peng *et al.* 2025), the practicalities of applying ARG inference to real data are understudied. In particular, there is a critical lack of software infrastructure to support evaluation and quality control of inferred ARGs.

Visualization is fundamentally important to data analysis. Many specialized tools exist to aid the visual analysis and quality control of genome assembly (Wick *et al.* 2015, Challis *et al.* 2020), read mapping (Robinson *et al.* 2011), and variant calling (Robinson *et al.* 2017, Tollefson *et al.* 2019, König *et al.* 2023), e.g. At every stage of a bioinformatics pipeline, it is important to visualize results to avoid artefacts and aid understanding of the data. Genome browsers such as IGV (Robinson *et al.* 2011) and the UCSC Genome Browser (Nassar *et al.* 2023) integrate many different data modalities, and are vital infrastructure for the field.

There is currently no straightforward means of visually summarizing ARGs, presenting a significant stumbling block for the nascent field of practical ARG inference. Inferred ARGs are essentially opaque, with only the most basic numerical summaries (such as numbers of nodes, mutations, etc.) or high-level statistics (Ralph *et al.* 2020) available. While tools for visualizing the local tree topologies exist, they are difficult to interpret and do not scale well to large sample sizes. To address this gap, we present `tsbrowse`, a client-server application providing genome browser-like functionality for ARGs. It provides interactive visualizations of the information structure of ARGs, smoothly scrolling from chromosome-level views down to individual nodes, edges and mutations. Supporting very large ARGs is a particular focus

Received: 28 April 2025; **Revised:** 26 June 2025; **Editorial Decision:** 28 June 2025; **Accepted:** 8 July 2025

for `tsbrowse`, as millions of genome sequences have been sampled for several species (Cesarani *et al.* 2022, Hunt *et al.* 2024, Stark *et al.* 2025) and ARGs of this scale are a particular focus of ongoing research (Kelleher *et al.* 2019, Anderson-Trocmé *et al.* 2023, Zhan *et al.* 2023, Zhang *et al.* 2023, Gunnarsson *et al.* 2024).

2 Results

2.1 Data model

`Tsbrowse` uses the 'succinct tree sequence' encoding of ARGs (Wong *et al.* 2024). This efficient ARG encoding is implemented by the `tskit` library (Ralph *et al.* 2020) and supported by most modern ARG simulation (Kelleher *et al.* 2016, 2018, Haller *et al.* 2019, Adrion *et al.* 2020, Baumdicker *et al.* 2022, Korfmann *et al.* 2023, Lauterbur *et al.* 2023, Tsambos *et al.* 2023, Tagami *et al.* 2024), inference (Kelleher *et al.* 2019, Speidel *et al.* 2019, Mahmoudi *et al.* 2022, Wohns *et al.* 2022, Zhan *et al.* 2023, Zhang *et al.* 2023, Deng *et al.* 2025), and processing methods (Fan *et al.* 2022, Nowbandegani *et al.* 2023). In `tskit`, ARGs are encoded as a collection of tables, storing information about the nodes and edges that describe the graph topology and the sites and mutations that encode the sequence variation (Fig. 1).

2.2 Architecture

`Tsbrowse` is a modular application written in Python, optimized for ARGs with millions of samples (Fig. 1). The application has two basic commands: `preprocess` and `serve`. The `preprocess` command takes an input ARG file and augments the `tskit` tables with additional columns, precomputing all the information required for visualization, and storing it as a compressed `.tsbrowse` file. To visualize an ARG we then run `tsbrowse serve`, which by default will open a browser window on the local machine, but also supports running the server on a remote machine across the network. This client-server architecture has some important advantages over a monolithic single-machine approach. Most importantly, the ARG being visualized remains in-situ and does not need to be downloaded from the server.

The goal of `tsbrowse` is to provide interpretable and interactive visual summaries of ARGs containing millions of nodes, edges and mutations. For instance, mutations have a clear interpretation when plotted on genome coordinate versus time axes, but at this scale the data density is far too high to simply plot each mutation as a point. We overcome this problem by using Datashader (Anaconda Developers and Contributors 2024) and the wider Holoviz ecosystem (Holoviz Developers 2024), which efficiently rasterizes large datasets at the requested resolution on the server, sends the image to the web browser for display, and dynamically updates as the user interactively navigates the ARG. This approach allows us to summarize very large ARGs interactively; Fig. 1, available as supplementary data at *Bioinformatics* online e.g. shows a screenshot of `tsbrowse` summarizing the 1.9 million mutations in a SARS-CoV-2 ARG with around 2.7 million nodes and edges.

2.3 User interface

The user interface is presented as a dashboard, with views to describe various aspects of the ARG. Figure 2 demonstrates the user interface using the Edges view as an example. The plot on the left is an overview of the 33 929 edges in a simulated ARG with a strong selective sweep (see Supplement for details). Each edge in the ARG is depicted as a horizontal line

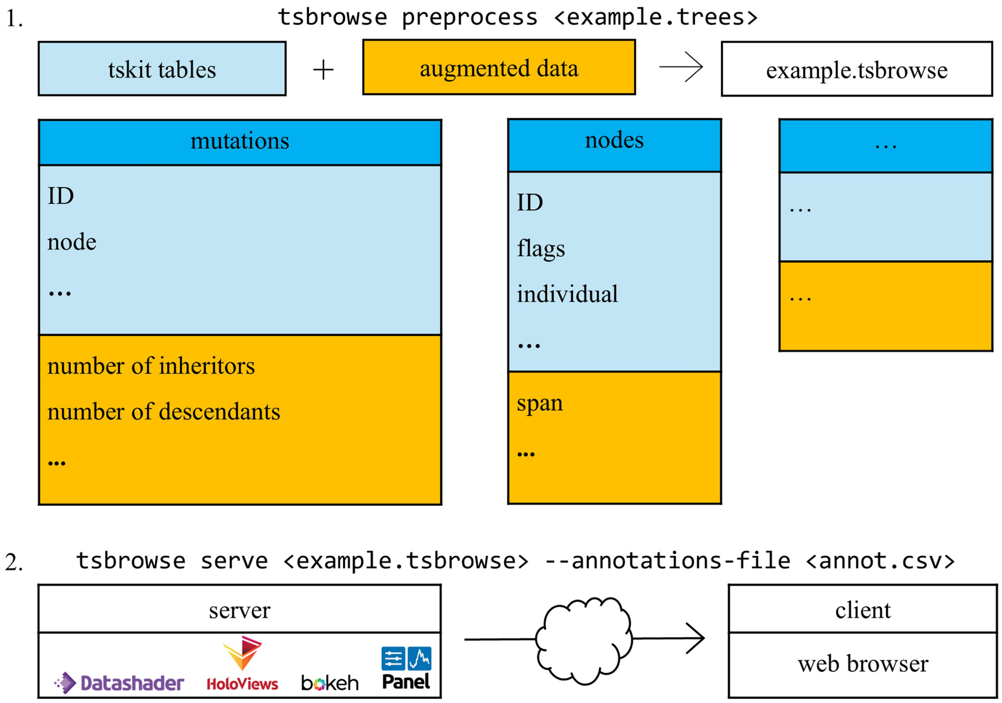
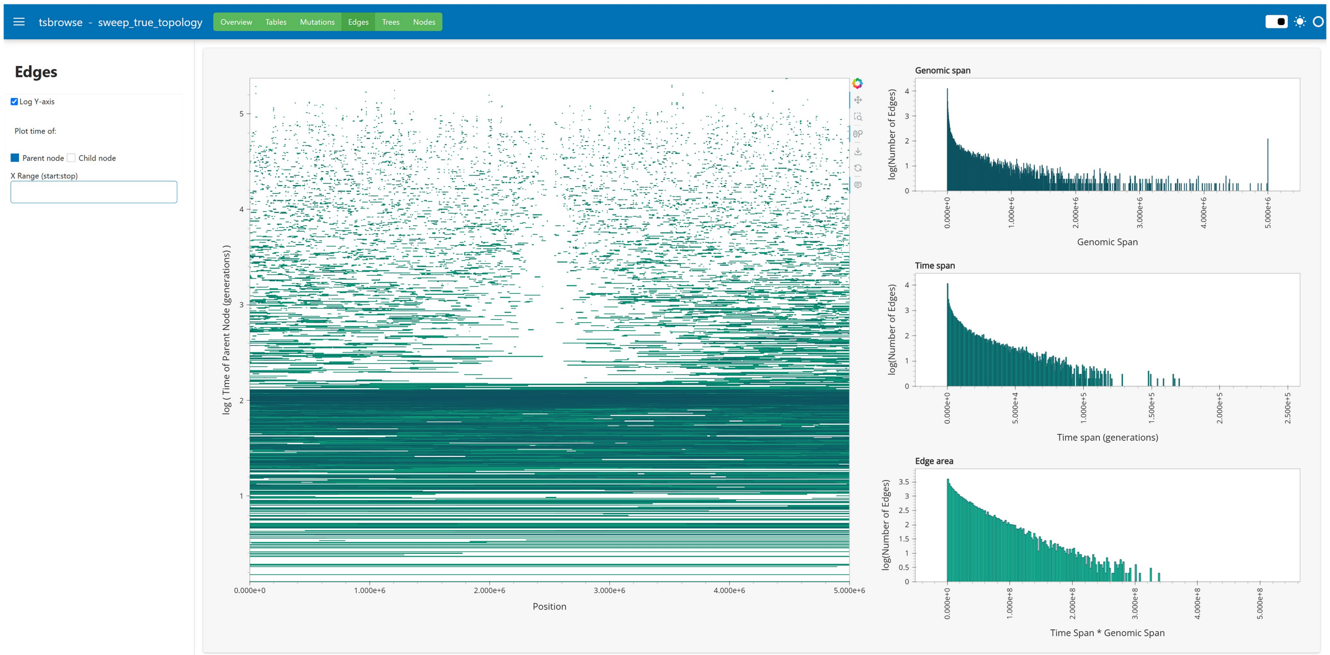


Figure 1. Overview of the tsbrowse architecture. Panel 1 depicts individual tskit tables, with table names in the headers and corresponding ARG properties listed beneath. In the pre-processing step, each table is augmented with additional information computed for visualization, shown as a separate section appended to the table. Panel 2 illustrates the serve step, which renders the output in a web browser using tools from the Holoviz ecosystem. Users may optionally supply an annotation file to overlay contextual information, such as overlapping genes or other sequence features.



tsbrowse - sweep_true_topology

Edges

☑ Log Y-axis

Plot time of:

■ Parent node ☐ Child node

X Range (start:stop)

Figure 2. Screenshot of the Edges view in `tsbrowse`. Visualization of the edges in a simulated ARG with a strong selective sweep in the middle of the genome (see Supplementary, available as supplementary data at *Bioinformatics* online for details). Edges are shown in the main browser pane on the left, with additional histograms summarizing the edges on the right. The effect of the sweep can be seen by the lack of edges crossing the centre of the simulated chromosome due to an excess of recent coalescent events. Moving away from the focal site, the oldest edges are the first to rejoin the ARG, followed by more recent edges, resulting in a wedge-like pattern of missing edges in the centre.

connecting the genomic coordinates of the parent and child nodes on the x-axis, and the y-axis shows the time of either the parent or the child, as chosen by the user. The user can interact with the main 'genome browser' window using a set of controls on the top-right corner provided by Bokeh (http://www.bokeh.pydata.org), allowing them to pan and zoom as required. The histograms to the right then summarize the edges as depicted in the browser window. A similar browser interface is provided for mutations (e.g. Fig. 1, available as supplementary data at *Bioinformatics* online) and nodes (e.g. Fig. 2, available as supplementary data at *Bioinformatics* online). `Tsbrowse` also provides an interactive table viewer with flexible searching and sorting utilities, which is a valuable debugging utility for developers.

2.4 Applications

The purpose of `tsbrowse` is to provide an interactive view of the `tskit` ARG data model to guide intuition, improve inference quality control and facilitate debugging. An example of how we can deepen our understanding using the genome-browser-like perspective provided by `tsbrowse` is given in the simulated ARG of Fig. 2, where the gap in the Edges view corresponds to the characteristic dip in diversity of a selective sweep. Comparing this ground truth to the ARGs inferred by four different inference methods in Fig. 3, available as supplementary data at *Bioinformatics* online, we can see that there are substantial qualitative differences between the results. These differences in ancestral haplotypes illustrated by the Edges view are unlikely to be captured by the tree-by-tree distance metrics usually used to evaluate inference methods (e.g. Kelleher *et al.* 2019, Zhang *et al.* 2023), providing further motivation for new and improved ways to compare simulated and inferred ARGs (Fritze *et al.* 2024).

Interest in ARG-based methods is burgeoning, but the methods are new, and practical guidance on applying inferences to real data is lacking. Data filtering is essential, and the effects of the choices that must be made along any bioinformatics pipeline on the final ARG are hard to predict and quantify. `Tsbrowse` was primarily developed as a way to quickly visualize the effects of such filtering choices on ARGs inferred by `tsinfer`, and it is now an indispensable element of the inference pipeline. Figure 4, available as supplementary data at *Bioinformatics* online shows a region of the 1000 Genomes data with gaps in site density that are spanned by exceptionally long edges, which are likely to bias downstream statistics. These interactive visualizations have also helped diagnose issues with the `tsinfer` inference algorithm. Figure 2, available as supplementary data at *Bioinformatics* online shows the genomic spans of ancestral nodes in an ARG inferred with `tsinfer`, demonstrating a clear excess of long haplotypes in the very ancient past. These insights have helped guide development and may lead to significant improvements in performance.

3 Discussion

Visualization of tree topology is a central task in phylogenetics, and although numerous tools exist (e.g. Huson *et al.* 2007, Vaughan 2017), the methods can typically only handle a few hundred nodes (but see e.g. Hadfield *et al.* 2018). Visualization of large-scale tree topologies with millions of nodes requires much more sophisticated approaches to capture topological features at different scales, and is an active research area (Wong and Rosindell 2022, Kramer *et al.* 2023). Adapting such methods, and integrating them into `tsbrowse` to provide a local tree viewer that operates at the million-node scale is an important direction for future work. An interactive viewer for the entire ARG topology, capturing semantic properties of the graph at a range of scales, is an even more ambitious goal, and would be a major asset for the field.

Acknowledgements

We are grateful to Yan Wong and the tskit community for helpful discussions and contributions.

Author contributions

Savita Karthikeyan (Conceptualization [equal], Investigation [equal], Methodology [equal], Software [equal], Validation [equal], Visualization [equal], Writing—original draft [equal], Writing—review & editing [equal]), Ben Jeffery (Software [equal], Visualization [equal]), Duncan Mbuli-Robertson (Methodology [equal], Software [equal], Visualization [equal], Writing—review & editing [equal]), and Jerome Kelleher (Conceptualization [equal], Methodology [equal], Project administration [equal], Software [equal], Supervision [equal], Writing—review & editing [equal])

Supplementary data

Supplementary data are available at *Bioinformatics* online.

Conflict of interest: None declared.

Funding

S.K. acknowledges full support and funding from Novo Nordisk Research Centre Oxford and the Biotechnology and Biological Sciences Research Council (UKRI-BBSRC) [BB/T008784/1]. D.M.-R. is funded by a studentship from the Wellcome Programme in Genomic Medicine and Statistics. J. K. acknowledges EPSRC [research grant EP/X024881/1], NIH [research grants HG011395 and HG012473], and the Robertson Foundation.

Data availability

The data underlying this article are available in https://github.com/savitakartik/tsbrowse-paper with DOI: https://doi.org/10.5281/zenodo.16615378

References

Adrion JR, Cole CC, Dukler N *et al.* A community-maintained standard library of population genetic models. *Elife* 2020;**9**:e54967.

Anaconda Developers and Contributors. 2024. Datashader. 10.5281/zenodo.15191180

Anderson-Trocmé L, Nelson D, Zabad S *et al.* On the genes, genealogies, and geographies of Quebec. *Science* 2023;**380**:849–55.

Baumdicker F, Bisschop G, Goldstein D *et al.* Efficient ancestry and mutation simulation with msprime 1.0. *Genetics* 2022;**220**:iyab229.

Brandt D, Wei X, Deng Y *et al.* Evaluation of methods for estimating coalescence times using ancestral recombination graphs. *Genetics* 2022;**221**:iyac044.

Brandt DY, Huber CD, Chiang CW *et al.* The promise of inferring the past using the ancestral recombination graph. *Genome Biol Evol* 2024;**16**:evae005.

Cesarani A, Lourenco D, Tsuruta S *et al.* Multibreed genomic evaluation for production traits of dairy cattle in the United States using single-step genomic best linear unbiased predictor. *J Dairy Sci* 2022;**105**:5141–52.

Challis R, Richards E, Rajan J *et al.* Blobtoolkit–interactive quality assessment of genome assemblies. *G3 (Bethesda)* 2020;**10**:1361–74.

Deng Y, Nielsen R, Song YS. Robust and accurate Bayesian inference of genome-wide genealogies for large samples. bioRxiv, https://doi.org/10.1101/2024.03.16.585351, 2024, preprint: not peer
reviewed.

Deng Y, Song YS, Nielsen R. A general framework for branch length estimation in ancestral recombination graphs. bioRxiv, https://doi.org/10.1101/2025.02.14.638385, 2025, preprint: not peer
reviewed.

Deraje P, Kitchens J, Coop G *et al.* Inferring the geographic history of recombinant lineages using the full ancestral recombination graph. bioRxiv, https://doi.org/10.1101/2024.04.10.588900, 2024, preprint: not peer reviewed.

Fan C, Mancuso N, Chiang CWK. A genealogical estimate of genetic relationships. *Am J Hum Genet* 2022;**109**:812–24.

Fritze H, Pope N, Kelleher J *et al.* A forest is more than its trees: haplotypes and inferred ARGs. bioRxiv, https://doi.org/10.1101/2024.11.30.626138, 2024, preprint: not peer reviewed.

Griffiths R, Marjoram P. An ancestral recombination graph. In: *Progress in Population Genetics and Human Evolution*. Springer. Conference date: 01-01-1997. 1997, 257–70.

Griffiths RC, Marjoram P. Ancestral inference from samples of DNA sequences with recombination. *J Comput Biol* 1996;**3**:479–502.

Grundler MC, Terhorst J, Bradburd GS. A geographic history of human genetic ancestry. *Science* 2025;**387**:1391–7.

Gunnarsson ÁF, Zhu J, Zhang BC *et al.* A scalable approach for genome-wide inference of ancestral recombination graphs. bioRxiv, https://doi.org/10.1101/2024.08.31.610248, 2024, preprint: not peer reviewed.

Hadfield J, Megill C, Bell SM *et al.* Nextstrain: real-time tracking of pathogen evolution. *Bioinformatics* 2018;**34**:4121–3.

Haller BC, Galloway J, Kelleher J *et al.* Tree-sequence recording in SLiM opens new horizons for forward-time simulation of whole genomes. *Mol Ecol Resour* 2019;**19**:552–66.

Hejase HA, Mo Z, Campagna L *et al.* A deep-learning approach for inference of selective sweeps from the ancestral recombination graph. *Mol Biol Evol* 2022;**39**:msab332.

Holoviz Developers. High-level tools to simplify visualization in Python. 2024. 10.5281/zenodo.3634719

Hudson RR. Properties of a neutral allele model with intragenic recombination. *Theor Popul Biol* 1983;**23**:183–201.

Hunt M, Hinrichs AS, Anderson D *et al.*; IMSSC2 Laboratory Network Consortium. Addressing pandemic-wide systematic errors in the sars-cov-2 phylogeny. bioRxiv, https://doi.org/10.1101/2024.04.29.591666, 2024, preprint: not peer reviewed.

Huson DH, Richter DC, Rausch C *et al.* Dendroscope: an interactive viewer for large phylogenetic trees. *BMC Bioinformatics* 2007;**8**:460–6.

Kelleher J, Etheridge AM, McVean G. Efficient coalescent simulation and genealogical analysis for large sample sizes. *PLoS Comput Biol* 2016;**12**:e1004842.

Kelleher J, Thornton KR, Ashander J *et al.* Efficient pedigree recording for fast population genetics simulation. *PLoS Comput Biol* 2018;**14**:e1006581.

Kelleher J, Wong Y, Wohns AW *et al.* Inferring whole-genome histories in large population datasets. *Nat Genet* 2019;**51**:1330–8.

König P, Beier S, Mascher M *et al.* DivBrowse–interactive visualization and exploratory data analysis of variant call matrices. *Gigascience* 2023;**12**:giad037.

Korfmann K, Abu Awad D, Tellier A. Weak seed banks influence the signature and detectability of selective sweeps. *J Evol Biol* 2023;**36**:1282–94.

Korfmann K, Sellinger TPP, Freund F *et al.* Simultaneous inference of past demography and selection from the ancestral recombination graph under the beta coalescent. *Peer Community J* 2024;**4**:e33.

Kramer AM, Sanderson T, Corbett-Detig R. Treenome Browser: co-visualization of enormous phylogenies and millions of genomes. *Bioinformatics* 2023;**39**:btac772.

Lauterbur ME, Cavassim MIA, Gladstein AL *et al.* Expanding the stdpopsim species catalog, and lessons learned for realistic genome simulations. *Elife* 2023;**12**:RP84874.

Lehmann B, Lee H, Anderson-Trocmé L *et al.* On ARGs, pedigrees, and genetic relatedness matrices. bioRxiv, https://doi.org/10.1101/2025.03.03.641310, 2025, preprint: not peer reviewed.

Lewanski AL, Grundler MC, Bradburd GS. The era of the ARG: an introduction to ancestral recombination graphs and their significance in empirical evolutionary genomics. *PLoS Genet* 2024;**20**:e1011110.

Link V, Schraiber JG, Fan C *et al.* Tree-based QTL mapping with expected local genetic relatedness matrices. *Am J Hum Genet* 2023;**110**:2077–91.

Mahmoudi A, Koskela J, Kelleher J *et al.* Bayesian inference of ancestral recombination graphs. *PLoS Comput Biol* 2022;**18**:e1009960.

Nassar LR, Barber GP, Benet-Pag S A *et al.* The UCSC genome browser database: 2023 update. *Nucleic Acids Res* 2023;**51**:D1188–95.

Nielsen R, Vaughn AH, Deng Y. Inference and applications of ancestral recombination graphs. *Nat Rev Genet* 2025;**26**:47–58.

Nowbandegani PS, Wohns AW, Ballard JL *et al.* Extremely sparse models of linkage disequilibrium in ancestrally diverse association studies. *Nat Genet* 2023;**55**:1494–502.

Osmond M, Coop G. Estimating dispersal rates and locating genetic ancestors with genome-wide genealogies. *Elife* 2024;**13**:e72177.

Pearson A, Durbin R. Local ancestry inference for complex population histories. bioRxiv, https://doi.org/10.1101/2023.03.06.529121, 2023, preprint: not peer reviewed.

Peng D, Mulder OJ, Edge MD. Evaluating ARG-estimation methods in the context of estimating population-mean polygenic score histories. *Genetics* 2025;**229**:iyaf033.

Ralph P, Thornton K, Kelleher J. Efficiently summarizing relationships in large samples: a general duality between statistics of genealogies and genomes. *Genetics* 2020;**215**:779–97.

Rasmussen MD, Hubisz MJ, Gronau I *et al.* Genome-wide inference of ancestral recombination graphs. *PLoS Genet* 2014;**10**:e1004342.

Robinson JT, Thorvaldsdóttir H, Winckler W *et al.* Integrative genomics viewer. *Nat Biotechnol* 2011;**29**:24–6.

Robinson JT, Thorvaldsdóttir H, Wenger AM *et al.* Variant review with the integrative genomics viewer. *Cancer Res* 2017;**77**:e31–e34.

Speidel L, Forest M, Shi S *et al.* A method for genome-wide genealogy estimation for thousands of samples. *Nat Genet* 2019;**51**:1321–9.

Stark Z, Glazer D, Hofmann O *et al.* A call to action to scale up research and clinical genomic data sharing. *Nat Rev Genet* 2025;**26**:141–7.

Stern AJ, Wilton PR, Nielsen R. An approximate full-likelihood method for inferring selection and allele frequency trajectories from DNA sequence data. *PLoS Genet* 2019;**15**:e1008384.

Tagami D, Bisschop G, Kelleher J. tstrait: a quantitative trait simulator for ancestral recombination graphs. *Bioinformatics* 2024;**40**:btae334.

Tollefson GA, Schuster J, Gelin F *et al.* VIVA (VIsualization of VAriants): a VCF file visualization tool. *Sci Rep* 2019;**9**:12648.

Tsambos G, Kelleher J, Ralph P *et al.* Link-ancestors: fast simulation of local ancestry with tree sequence software. *Bioinform Adv* 2023;**3**:vbad163.

Vaughan TG. IcyTree: rapid browser-based visualization for phylogenetic trees and networks. *Bioinformatics* 2017;**33**:2392–4.

Whitehouse LS, Ray DD, Schrider DR. Tree sequences as a general-purpose tool for population genetic inference. *Mol Biol Evol* 2024;**41**:msae223.

Wick RR, Schultz MB, Zobel J *et al.* Bandage: interactive visualization of de novo genome assemblies. *Bioinformatics* 2015;**31**:3350–2.

Wohns AW, Wong Y, Jeffery B *et al.* A unified genealogy of modern and ancient genomes. *Science* 2022;**375**:eabi8264.

Wong Y, Rosindell J. Dynamic visualisation of million-tip trees: the OneZoom project. *Methods Ecol Evol* 2022;**13**:303–13.

Wong Y, Ignatieva A, Koskela J *et al.* A general and efficient representation of ancestral recombination graphs. *Genetics* 2024;**228**:iyae100.

Zhan SH, Ignatieva A, Wong Y *et al.* Towards pandemic-scale ancestral recombination graphs of sars-cov-2. bioRxiv, https://doi.org/10.1101/2023.06.08.544212, 2023, preprint: not peer reviewed.

Zhang BC, Biddanda A, Gunnarsson ÁF *et al.* Biobank-scale inference of ancestral recombination graphs enables genealogical analysis of complex traits. *Nat Genet* 2023;**55**:768–76.

Bioinformatics, 2025, 41, 1–5
https://doi.org/10.1093/bioinformatics/btaf393
Applications Note