Correspondence

The SplitsTree App: interactive analysis and visualization using phylogenetic trees and networks

Since Darwin, phylogenetic trees have been widely used to represent the evolutionary history of species and other taxonomic groups. In cases where reticulate evolutionary events, such as horizontal gene transfer, speciation by hybridization or reassortment, play a notable role, phylogenetic networks are sometimes employed. Here, we present the SplitsTree App, which is software for exploring and analyzing phylogenetic data using both trees and networks. Offering a comprehensive set of features, the application provides over 100 algorithms for computing distances, phylogenetic trees, split networks, haplotype networks, rooted phylogenetic networks, tanglegrams, consensus trees and consensus networks. It introduces a faster version of neighbor-net[1] and provides an implementation of median-joining, which are the two most widely used techniques for constructing phylogenetic networks. Moreover, the software incorporates genetic and phylogenetic indices, ranging from Tajima's D for sequences to Shapely values on phylogenetic networks. A complete description of all the program's features can be found in the Supplementary User Manual.
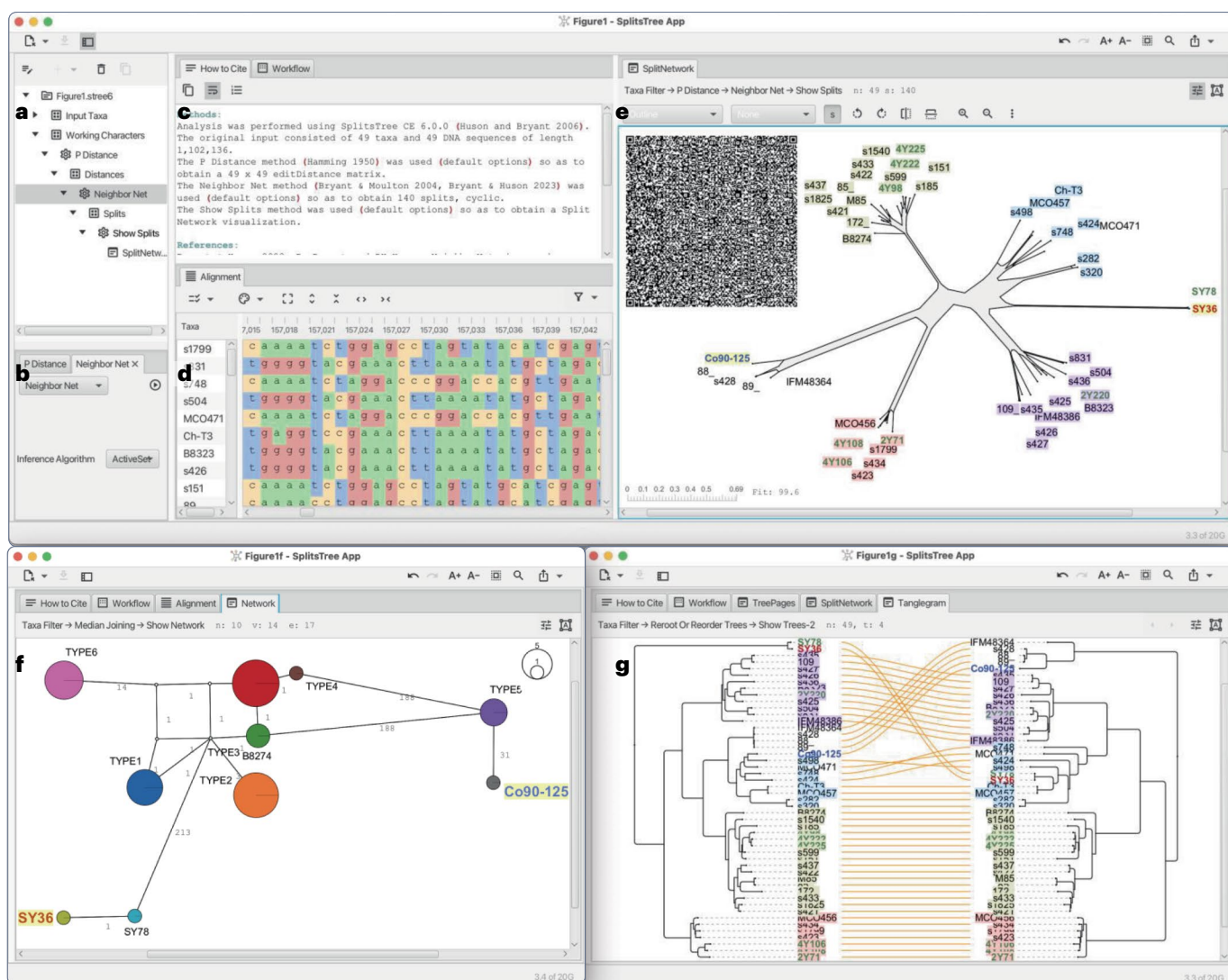
Correspondence

Fig. 1 | In the new SplitsTree App, each dataset is represented by a single window and different analyses of the dataset appear as different tabs in that window. a, The sidebar gives access to the current workflow. **b**, These tabs allow the setting of parameters and execution of algorithms. **c**, The "how to cite" tab provides a summary of methods and relevant references. **d**, The alignment viewer, here loaded with an alignment of 1.1 million SNPs across 49 fungal isolates. **e**, A phylogenetic outline computed from the alignment, with a QR code containing a digital description that can be scanned and opened in SplitsTree App or saved as text. **f**, A median-joining network illustrating relationships between different haplotypes. **g**, A tanglegram of two trees obtained by data aligned to two different reference genomes. Based on fungal sequencing data[6].

This new software is designed to accommodate the increasing scale and intricacy of modern datasets. It extends, integrates and supersedes our earlier applications SplitsTree4 (ref. [2]) for unrooted phylogenetic trees and networks, Dendroscope3 (ref. [3]) for rooted trees and networks, and PopArt[4] for haplotype analysis. Unrooted networks are widely used in phylogenetics and population genetics[5,6], but also in other fields such as linguistics and anthropology[7]. While rooted phylogenetic networks have the potential to explicitly represent reticulate evolution[8], they have faced limited adoption because of a lack of applicable software. To address this, the SplitsTree App provides new methods for computing these networks[9].

The application is designed as a versatile tool for exploring data from many sources. It accepts input in several formats, including sequences, alignments, distances, phylogenetic trees and networks, accommodating a wide range of data types. The software facilitates the assessment of sequence data quality and properties, allowing researchers to identify conflicting signals resulting from reticulate evolution, incomplete lineage sorting, polymorphic data or labeling errors. This critical evaluation should be done before conducting any model-based analysis that assumes a single tree structure a priori. Another key application is the use of phylogenetic network methods with datasets for which reticulate evolutionary events — horizontal gene transfer or hybridization, or borrowing in linguistics — are suspected to have played a role.

Figure 1 shows the user interface, illustrated using single-nucleotide polymorphism (SNP) data from a recent study on the evolution of pathogenic fungi from hybrid isolates collected from warm marine environments[6].

Reproducibility, reusability and authorship attribution are fundamental principles in scientific software development. The SplitsTree App embodies these principles by adopting a workflow-driven approach, wherein all data and analysis steps are explicitly represented in a provenance graph. This design enables users to save and later restore all intermediate results of an analysis, ensuring transparency and reproducibility. Users can interactively set up a workflow, export it to a file and execute it on multiple datasets from the command line. With the aid of the workflow, the program maintains a human-readable summary that includes data, algorithms, parameters and references, facilitating result interpretation and proper attribution. In biology publications, phylogenetic trees and networks are often depicted only as images, lacking accessible data. To address this, the SplitsTree App can embed a QR code within the image. This provides a digital description of the phylogenetic tree or network in Newick format, improving data sharing and reuse. One limitation of the software is its lack of algorithms for model-based inference of phylogenetic networks. Developing such methods is a challenging task and remains an active area of research.

Code availability
The Java source code and application installers for Linux, MacOS and Windows are available at https://github.com/husonlab/splitstree6 (GPLv3 license).

Daniel H. Huson [iD] [1,2] ✉ **& David Bryant** [iD] [3]

[1]Institute for Bioinformatics and Medical Informatics, University of Tübingen, Tübingen, Germany. [2]Excellence Cluster "Controlling Microbes to Fight Infection", University of Tübingen, Tübingen, Germany. [3]David Bryant, Department of Mathematics and Statistics, University of Otago, Dunedin, New Zealand.
✉e-mail: daniel.huson@uni-tuebingen.de

References
1. Bryant, D. & Huson, D. H. *Front. Bioinform.* **3**, ••• (2023).
2. Huson, D. H. & Bryant, D. *Mol. Biol. Evol.* **23**, 254–267 (2006).
3. Huson, D. H. & Scornavacca, C. *Syst. Biol.* **61**, 1061–1067 (2012).
4. Leigh, J. W. & Bryant, D. *Methods Ecol. Evol.* **6**, 1110–1116 (2015).
5. Molteni, C., Forni, D., Cagliani, R., Clerici, M. & Sironi, M. *Vaccines (Basel)* **7**, 92 (2022).
6. Del Olmo, V. et al. *Nat. Commun.* **14**, 6919 (2023).
7. Fulwood, E. L. *Humanit. Soc. Sci. Commun.* **9**, 392 (2022).
8. Koblmüller, S. et al. *BMC Evol. Biol.* **7**, 7 (2007).
9. Zhang, L., Abhari, N., Colijn, C. & Wu, Y. *Genome Res.* **33**, 1053–1060 (2023).

Author contributions
D.H.H. contributed the main software framework and both authors contributed implementations of algorithms. Both authors contributed equally to the manuscript and user manual.

Competing interests
The authors declare no competing financial interests

Additional information
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41592-024-02406-3.

Peer review information *Nature Methods* thanks Huw Ogilvie, Kristina Wicke and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.