# ggret: An R package for visualising and manipulating tree-based phylogenetic networks

Gerd Specht [1,2¶], **Clemens Schmid** [3], **Denise Kühnert** [2,4], **and Arthur Kocher** [2,3¶]

1 Faculty of Mathematics and Computer Science, Friedrich Schiller University Jena, Germany **2** Transmission, Infection, Diversification & Evolution Group, Max Planck Institute of Geoanthropology, Jena, Germany **3** Department of Archaeogenetics, Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany **4** Centre for Artificial Intelligence in Public Health Research, Robert Koch Institute, Wildau, Germany **¶** Corresponding author

Summary

Evolutionary relationships of biological entities are most often modeled with phylogenetic trees. Phylogenetic trees consist of branches (or edges) representing direct lines of descent or genetic flow from ancestor to offspring (*i.e.* lineages), and nodes representing evolutionary "splits" through which a parental lineage gives rise to multiple child lineages. This vertical model of evolution has provided immense insights into the evolutionary history and processes underlying observed biological diversity. However, it fails to account for "horizontal" modes of evolution, whereby genetic material can be exchanged between contemporaneous organisms through a variety of mechanisms across the tree of life (Arnold et al., 2022; Keeling, 2024; Pérez-Losada et al., 2015). In recent years, advances in sequencing technologies and computational methods have made it increasingly possible to integrate horizontal evolutionary events into reticulated phylogenetic trees (or phylogenetic networks; (Chen & Wang, 2010; Huson & Bryant, 2006; Müller et al., 2020, 2022; Vaughan et al., 2017)). While phylogenetic networks have the potential to provide more comprehensive and accurate evolutionary pictures for many biological groups, the development of specific tools is required for their manipulation and visualisation. Here, we present ggret, an R package building upon the popular ggtree package (Yu et al., 2017) for the manipulation and visualisation of phylogenetic networks. ggret provides novel functions for parsing extended Newick and NEXUS files and introduces the geom_ret object for visualisation of tree-based phylogenetic networks.

Statement of need

The R language is commonly used for phylogenetic analysis and visualisation with packages such as ape and ggtree providing important functionalities for handling and visualizing classic phylogenetic trees (Paradis & Schliep, 2019; Yu et al., 2017). In contrast, R packages for the manipulation and plotting of phylogenetic networks are largely lacking. The ape and phangorn packages (Paradis & Schliep, 2019; Schliep, 2011) provide basic plotting functions for explicit networks and split networks. These have been extended in the tanggle package to allow more flexibility using the ggplot2 syntax (Wickham, 2016), building upon ggtree. The ggret package presented here complements these existing tools by providing alternative modes of visualization of explicit phylogenetic networks as well as additional functionalities. In particular, ggret enables the parsing of phylogenetic networks in NEXUS formats such as those produced by the popular *BEAST2* software (Bouckaert, 2019), and can handle associated metadata via treedata objects, allowing to leverage the versatile ggtree annotation and visualisation methods.

Usage

ggret is available on GitHub. All instructions for installation and usage can be found in the package vignette and website.

Below are some examples of phylogenetic network visualization with ggret (Figure 1). The phylogenetic network was simulated with the *BEAST2* package *Bacter* (Vaughan et al., 2017) and parsed together with node metadata using the read_beast_retnet function. The resulting retnet treedata object has been included in the package for the sake of reproducibility.
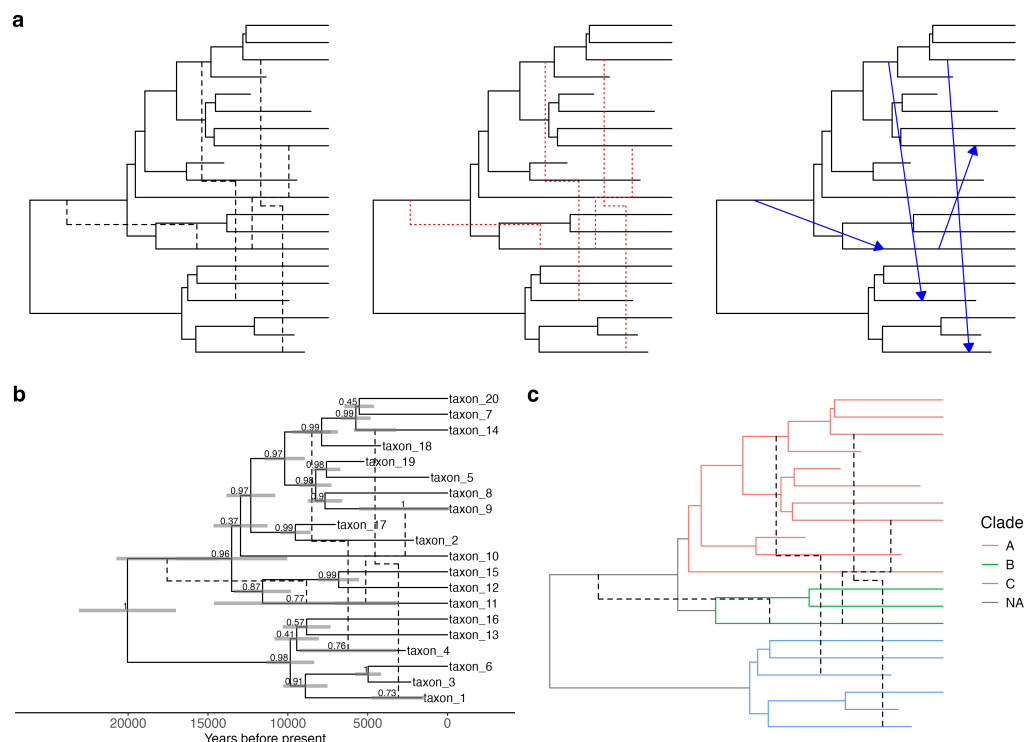
Figure 1: Examples of phylogenetic network visualization with *ggret*. **a.** Reticulation edges can be plotted in different shapes and colors. **b.** Annotated phylogenetic network. Tip labels are plotted and internal nodes are labelled with their posterior probability. The x-axis represents a timeline in years before present. The 95% highest probability density intervals of nodes' ages are indicated by grey bars. **c.** Phylogenetic network with colored based on clade information. Deep branches not belonging to any defined clade are labelled as NA

Acknowledgements

References

Arnold, B. J., Huang, I.-T., & Hanage, W. P. (2022). Horizontal gene transfer and adaptive evolution in bacteria. *Nature Reviews Microbiology*, *20*(4), 206–218. https://doi.org/10.1038/s41579-021-00650-4

Bouckaert, T. G. A. B.-S., Remco AND Vaughan. (2019). BEAST 2.5: An advanced software platform for bayesian evolutionary analysis. *PLOS Computational Biology*, *15*(4), 1–28. https://doi.org/10.1371/journal.pcbi.1006650

Chen, Z.-Z., & Wang, L. (2010). HybridNET: A tool for constructing hybridization networks. *Bioinformatics (Oxford, England)*, *26*(22), 2912–2913. https://doi.org/10.1093/bioinformatics/btq548

Huson, D. H., & Bryant, D. (2006). Application of phylogenetic networks in evolutionary studies. *Molecular Biology and Evolution*, *23*(2), 254–267. https://doi.org/10.1093/molbev/msj030

Keeling, P. J. (2024). Horizontal gene transfer in eukaryotes: Aligning theory with data. *Nature Reviews Genetics*, 1–15. https://doi.org/10.1038/s41576-023-00688-5

Müller, N. F., Kistler, K. E., & Bedford, T. (2022). A bayesian approach to infer recombination patterns in coronaviruses. *Nature Communications*, *13*(1), 4186. https://doi.org/10.1038/s41467-022-31749-8

Müller, N. F., Stolz, U., Dudas, G., Stadler, T., & Vaughan, T. G. (2020). Bayesian inference of reassortment networks reveals fitness benefits of reassortment in human influenza viruses. *Proceedings of the National Academy of Sciences*, *117*(29), 17104–17111. https://doi.org/10.1073/pnas.1918304117

Paradis, E., & Schliep, K. (2019). Ape 5.0: An environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics*, *35*(3), 526–528. https://doi.org/10.1093/bioinformatics/bty633

Pérez-Losada, M., Arenas, M., Galán, J. C., Palero, F., & González-Candelas, F. (2015). Recombination in viruses: Mechanisms, methods of study, and evolutionary consequences. *Infection, Genetics and Evolution*, *30*, 296–307. https://doi.org/10.1016/j.meegid.2014.12.022

Schliep, K. P. (2011). Phangorn: Phylogenetic analysis in r. *Bioinformatics*, *27*(4), 592–593. https://doi.org/10.1093/bioinformatics/btq706

Vaughan, T. G., Welch, D., Drummond, A. J., Biggs, P. J., George, T., & French, N. P. (2017). Inferring ancestral recombination graphs from bacterial genomic data. *Genetics*, *205*(2), 857–870. https://doi.org/10.1534/genetics.116.193425

Wickham, H. (2016). *ggplot2 Elegant Graphics for Data Analysis*. Springer International Publishing. https://doi.org/10.1007/978-3-319-24277-4

Yu, G., Smith, D. K., Zhu, H., Guan, Y., & Lam, T. T.-Y. (2017). Ggtree: An R package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods in Ecology and Evolution*, *8*(1), 28–36. https://doi.org/10.1111/2041-210X.12628