

RESEARCH ARTICLE

Topo-phylogeny: Visualizing evolutionary relationships on a topographic landscape

Jamie Waese^{1*}, Nicholas J. Provart^{1,2}, David S. Guttman^{1,2}

1 Department of Cell & Systems Biology, University of Toronto, Toronto, Ontario, **2** Centre for the Analysis of Genome Evolution & Function, University of Toronto, Toronto, Ontario

* jamie.waese@mail.utoronto.ca

Abstract

Phylogenetic trees are the de facto standard for visualizing evolutionary relationships, but large trees can be difficult to interpret because they require a high cognitive load to identify relationships between multiple operational taxonomic units (OTUs). We present a new tool for displaying phylogenetic relationships as a topographic map in which OTUs autonomously attract or repel one another based on their individual branch lengths and distance to a common ancestor. This data visualization paradigm makes it possible to preattentively identify the nature of the relationship between items without having to trace a complex network of branches back to the root. This tool was developed for exploring phylogenetic data, but the technique could be extended for visualizing other hierarchical structures as well.



OPEN ACCESS

Citation: Waese J, Provart NJ, Guttman DS (2017) Topo-phylogeny: Visualizing evolutionary relationships on a topographic landscape. PLoS ONE 12(5): e0175895. <https://doi.org/10.1371/journal.pone.0175895>

Editor: Robert Guralnick, University of Colorado, UNITED STATES

Received: September 20, 2016

Accepted: April 2, 2017

Published: May 1, 2017

Copyright: © 2017 Waese et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: The source code can be downloaded from Figshare at the following link: <https://figshare.com/s/15ce7763fcd97d778752>.

Funding: The authors received no specific funding for this work.

Competing interests: The authors have declared that no competing interests exist.

1 Introduction

Phylogenetic trees and cluster diagrams are essential tools for displaying evolutionary and similarity relationships among species, gene or protein sequences, and many other types of data. These tree-like branching diagrams are generally intuitive and easy to read, but they do not scale well when working with very large numbers of operational taxonomic units (OTUs). Problems arise due to a poor overall use of space, arbitrary clade ordering, hard to read labels (especially with radial trees), and the high cognitive load required for tracing lineages to their common ancestral nodes to determine lineage and clade relationships. While these are trivial issues when working with most moderately size trees, they can become problematic when the trees become very large, or with trees that show early diversification and long terminal branches. Despite these problems, phylogenetic trees remain the de facto standard for displaying evolutionary relationships, despite very little work exploring the effects of this visualization approach on understanding [1].

Several tools have attempted to address these issues, for example PhyloMap [2], TreeJuxtaposer [3], TreeWiz [4], Dendroscope [5], Treevolution [6], PAUP [7], Mesquite [8], ToLWeb [9], FigTree [10], PhyloWidget [11], OneZoom [12] and others. An impressive list of 392 different phylogeny visualization and analysis packages is maintained by Felsenstein [13]. These tools enable users to zoom in and out of branches, compress clades into groups and produce circular trees and other layouts that display a large amount of information in a limited space. Hughes et al [14] developed a unique method for displaying large phylogenetic trees in 3D

hyperbolic space, taking advantage of an additional dimension to visualize and navigate more than 100,000 nodes. However each of these approaches maintains the underlying tree metaphor that requires the viewer to use “top down” cognitive visual processes [15] to determine the nature of a relationship between two OTUs. Simply put, trees require conscious direction of the gaze as it travels up and down the network to determine how two OTUs are linked.

Trees are not the only way to visualize phylogenetic relationships. Isabel Meirelles [16] categorizes visual depictions of hierarchical structures into two basic graphical forms: stacked and nested schemes. Whereas phylogenetic trees are stacked schemes with lines connecting the elements in their set, nested schemes depict elements as containers that are grouped and assembled according to their hierarchical relationships. One of the best known nested schemes is Ben Schneiderman’s “treemap” [17] which displays hierarchical data as a set of nested rectangles with areas that are proportional to a specified dimension of the data. There are several variations of treemap layouts: icicle, squarified, pivot by split size, Voronoi, radial icicle [16], and cascaded [18]. Each of these methods offers certain advantages over stacked schemes such as: (i) they use screen real estate more efficiently, (ii) they make it easier to recognize groups, patterns and outliers across large data sets, and (iii) they can convey multiple dimensions of data through size, colour, border quality and shape. While treemaps are useful for representing data categories that differ in their relative abundance, they can be difficult to interpret since the rectangular bins representing category abundance are typically of different shapes, and the level of bin nesting is often difficult to observe. Nevertheless, one treemap approach that has been successfully applied to phylogenetic visualization is MetaTreeMap [19], which uses the D3.js library [20].

We decided to explore a different approach to visualizing hierarchical information inspired by the work of Max Fürbringer, who in 1888 drew a phylogenetic tree of birds that included a series of “slices” from the middle of the tree [21] (Fig 1A & 1B). This graphic resembles a nested scheme called “circle packing” [22] however the bounding shapes are irregular, allowing for a much more efficient use of space. The visualization approach is also reminiscent of that used in topographic maps (Fig 1C), where contour lines are used to indicate common levels of elevation, and consequently, higher levels of elevation are nested within lower levels. When applied in a phylogenetic context, phylogenetic depth (time to most recent common

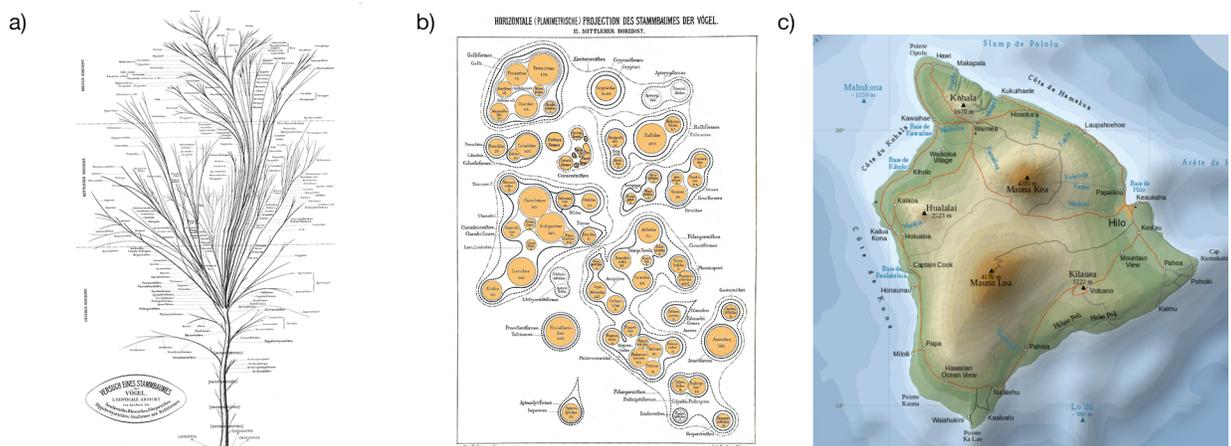


Fig 1. a) Max Fürbringer’s “Phylogenetic Tree of Birds”. b) A horizontal projection from the middle section of the tree. These figures, reproduced from https://en.wikipedia.org/wiki/Max_F%C3%BCrbringer, were originally published in 1888 and are in the public domain. c) A topographic map of Hawaii island, retrieved from [https://en.wikipedia.org/wiki/Hawaii_\(island\)#/media/File:Hawaii_Island_topographic_map-en.svg](https://en.wikipedia.org/wiki/Hawaii_(island)#/media/File:Hawaii_Island_topographic_map-en.svg). Reprinted with permission as per <http://creativecommons.org/licenses/by-sa/3.0/>.

<https://doi.org/10.1371/journal.pone.0175895.g001>

ancestry) is analogous to geospatial elevation, contour lines indicate the nested clustering of lineages and clades, and gaps between clusters indicate evolutionarily distinct clades.

The idea of using topographic maps to visualize hierarchical information has been explored by others. La Rosa et. al [23] demonstrate a method to create a topographic representation of bacteria clusters organized in a rectangular lattice that defines data neighborhood relationships. Unfortunately, the charts produced are non-intuitive and consequently difficult to decipher. A project by Cortese et. al [24] explores the use of topographic maps to visualize the connections between internet service providers and track the paths of information packets across a hierarchical network. Their overall approach is very appealing, but their topographic clustering is inverted with root and low level nodes getting drawn as high altitude peaks, while terminal and high level nodes are drawn at lower altitudes around the fringes.

2 Implementation

We present a new tool for displaying phylogenetic relationships as a topographic map in which OTUs autonomously attract or repel one another based on their individual branch lengths and distance to a common ancestor to form “geographic regions” of related “peaks” (Fig 2). Topographic contour shapes indicate the level at which different OTUs are connected. The width of the contour around any given OTU indicates the branch lengths of each of its ancestors going back to the root. OTUs enclosed by the same contour shape share an ancestor. Valleys between the various plateaus and peaks indicate clades from a different lineage. We call this data visualization paradigm a “topo-phylogeny”.

The program accepts data in Newick format, and can either be pasted into an input box or loaded from a file. All data processing is done within the browser on the client side. The Newick file gets parsed with Newick.js [25] and converted into a JSON object. This object is passed into d3.phylogram.js [26] to build a standard phylogenetic tree visible on the left side of the page.

On the right side of the page, the topo-phylogeny chart is created as an SVG image. Each node consists of a stack of circles, with the top circle containing the label and each circle below it sized according to the cumulative branch length at any given depth. Circle sizes are determined

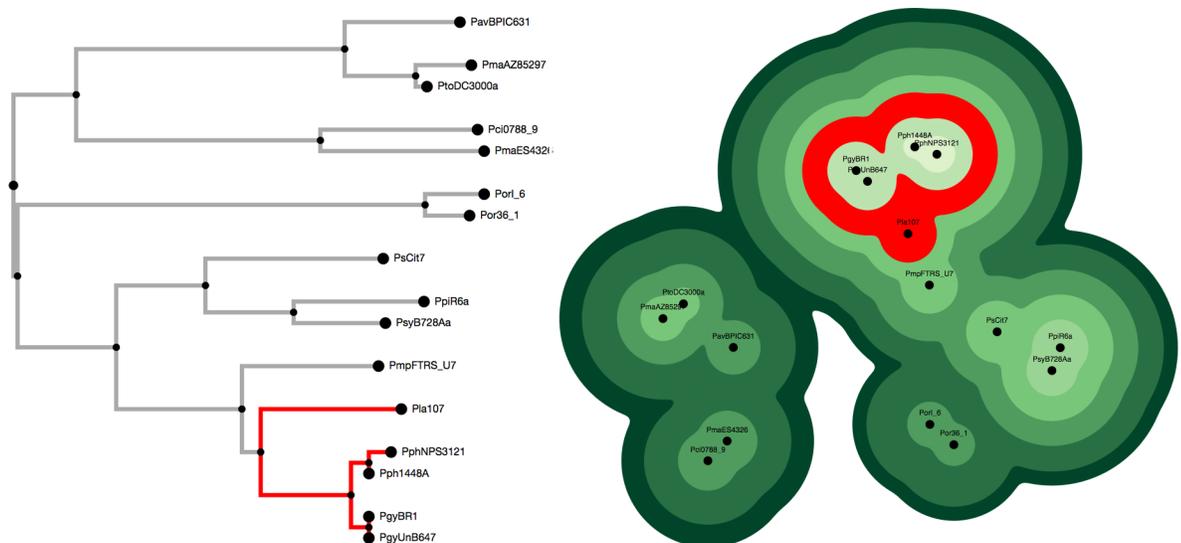


Fig 2. a) Phylogenetic tree with 16 OTUs. b) Topo-phylogeny chart based on the same structure. The equivalent branches are highlighted red in both figures.

<https://doi.org/10.1371/journal.pone.0175895.g002>

3 Results

Topo-phylogeny charts with fewer than 20 OTUs take less than ten seconds to generate on a 2015 model 2.2 GHz Intel Core i7 MacBook Air with 8GB of RAM. Topo-phylogeny charts with more than a hundred OTUs can take several minutes before nodes settle into a stable position, however the major groupings are usually set quickly. Some topo-phylogeny charts may remain in constant motion if there is no ideal position for one or more of the OTUs. A 'Pause' button toggles the force layout function off and on if the user wishes to freeze the display. Hovering over the OTUs in either chart will highlight the equivalent OTU on both charts. This makes it possible to shift one's gaze back and forth between both visualization methods without getting disoriented. Items of interest can be tagged with red or green markers by clicking or shift-clicking on them.

The technique scales well up to several hundred OTUs. Whereas large phylogenetic trees can be difficult to display and read, the underlying structure of a topo-phylogeny remains apparent (Fig 4A & 4B). Larger maps may have poor user interface response times because each node continuously compares its position against every other node, occupying system resources. Hug et al.'s Tree of Life [27] with 3740 OTUs and a maximum depth of 72 takes twenty minutes for the nodes to load, get drawn to the screen and find stable positions (Fig 4C), but the program is virtually unresponsive during this time. The program includes an option to temporarily pause screen updates so the nodes can adjust their positions in the background without having to reposition all the DOM elements for each cycle. For large data sets, there is an option to adjust the "tightness" of the layout by changing a constant value (i.e.,

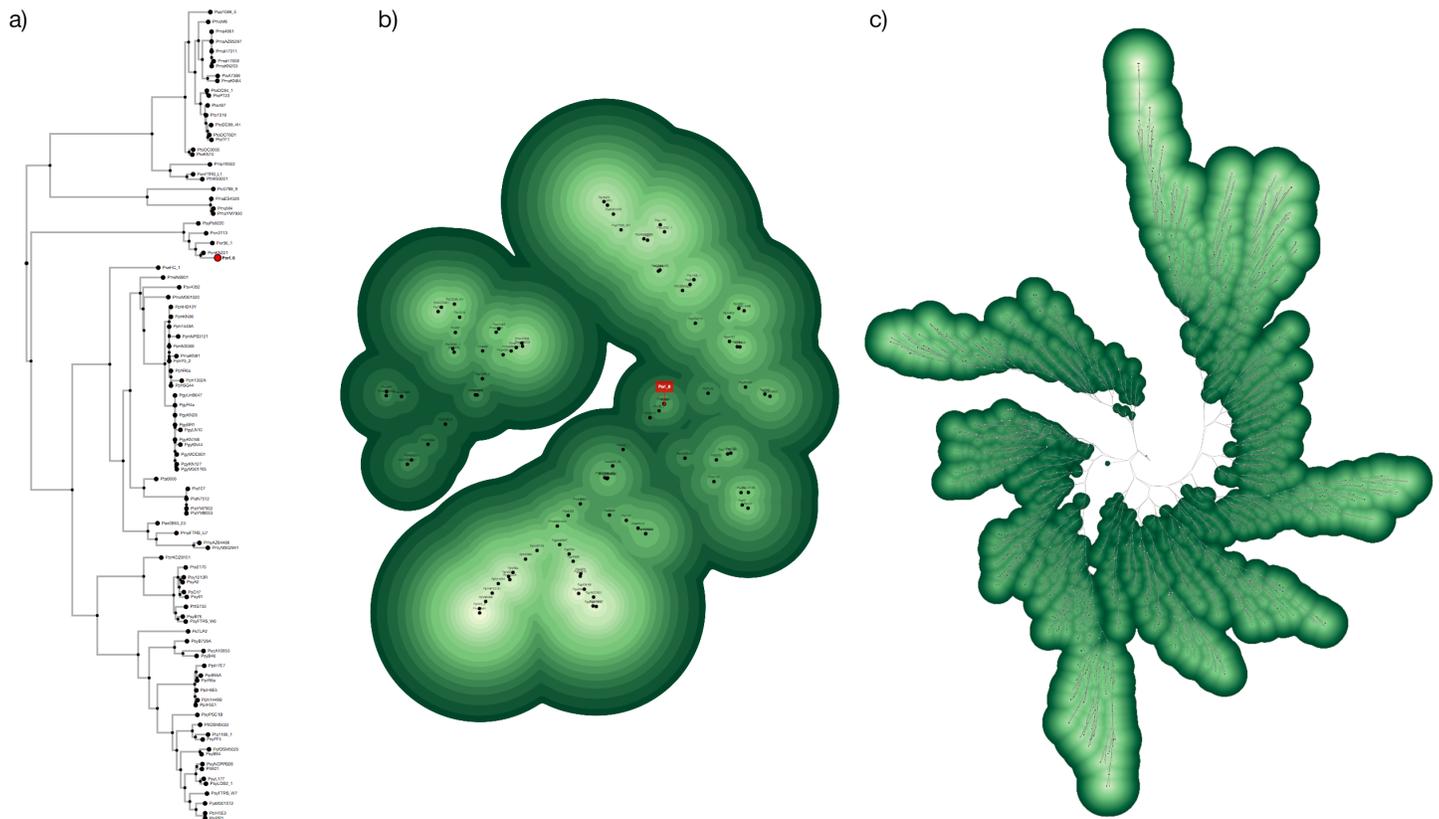


Fig 4. a) Phylogenetic tree with 96 OTUs. b) Topo-phylogeny chart based on the same structure. The same OTU is highlighted red in both figures. c) Topo-phylogeny chart based on Hug et al.'s Tree of Life (2016) with 3740 OTUs.

<https://doi.org/10.1371/journal.pone.0175895.g004>

radius length) that is used to calculate the initial positions of each node. Future versions of the tool will focus on improving interactivity for very large data files.

4 Discussion

We present a new data visualization paradigm for exploring phylogenetic data. While topographic maps are a familiar method for displaying geospatial elevation data, their underlying visual paradigm can be equally well applied to other kinds of hierarchical data structures. Phylogenetic data are well suited to a topographic layout, and the use of topo-phylogeny charts can complement traditional phylogenetic approaches. Most notably, topo-phylogeny charts eliminate vertical scanning bias, where the proximity of OTUs at the tips of a tree is often confounded with their evolutionary relatedness—a mistake that is commonly made by individuals with limited training in phylogenetics. The free movement of nested clades in the two dimensional topo-phylogenetic space more accurately reflects the ability of nodes to freely rotate, thereby changing the relative position of OTUs, while retaining their nested evolutionary relationships.

Topo-phylogeny charts take advantage of preattentive visual processing [28] to help viewers grasp hierarchical relationships between items with a single glance. The end product looks like a topographic map in which each element's position is determined by on its own attractions, repulsions and individual size. This visualization approach makes it easy to identify related and unrelated items because their group status preattentively “pops out” to the viewer. Items that are clustered together and surrounded by the same colour of contour shape are closely related; items that are in separate clusters and appear to have deep valleys between them are not. These relationships can be identified at a glance without having to trace individual branches back to the root.

Topo-phylogram is not intended as a replacement for the phylogenetic tree. Rather, we propose this data visualization paradigm as a complementary view. Viewing phylogenetic trees and topo-phylogeny charts side by side could improve viewers understanding of hierarchical structures because each method takes advantage of different visual processes.

Acknowledgments

We thank Wolfgang Stuerzlinger, Asher Pasha and Alex Waese-Perlman for helpful discussion and advice.

Author Contributions

Conceptualization: DG JW NP.

Funding acquisition: DG.

Methodology: JW.

Project administration: DG.

Software: JW.

Supervision: NP.

Validation: DG NP.

Visualization: JW.

Writing – original draft: JW.

Writing – review & editing: DG NP.

References

1. Carrizo SF. Phylogenetic Trees: An Information Visualisation Perspective. In: Proceedings of the Second Conference on Asia-Pacific Bioinformatics—Volume 29 [Internet]. Darlinghurst, Australia, Australia: Australian Computer Society, Inc.; 2004 [cited 2014 Oct 27]. p. 315–320. (APBC '04). Available from: <http://dl.acm.org/citation.cfm?id=976520.976563>
2. Zhang J, Mamlouk AM, Martinetz T, Chang S, Wang J, Hilgenfeld R. PhyloMap: an algorithm for visualizing relationships of large sequence data sets and its application to the influenza A virus genome. *BMC Bioinformatics*. 2011 Jun 20; 12(1):248.
3. Munzner T, Guimbretière F, Tasiran S, Zhang L, Zhou Y. TreeJuxtaposer: Scalable Tree Comparison Using Focus+Context with Guaranteed Visibility. In: ACM SIGGRAPH 2003 Papers [Internet]. New York, NY, USA: ACM; 2003 [cited 2014 Oct 29]. p. 453–462. (SIGGRAPH '03).
4. Rost U, Bornberg-Bauer E. TreeWiz: interactive exploration of huge trees. *Bioinformatics*. 2002 Jan 1; 18(1):109–14. PMID: [11836218](https://pubmed.ncbi.nlm.nih.gov/11836218/)
5. Huson DH, Richter DC, Rausch C, Dezulian T, Franz M, Rupp R. Dendroscope: An interactive viewer for large phylogenetic trees. *BMC Bioinformatics*. 2007; 8:460. <https://doi.org/10.1186/1471-2105-8-460> PMID: [18034891](https://pubmed.ncbi.nlm.nih.gov/18034891/)
6. Santamaría R, Therón R. Treevolution: visual analysis of phylogenetic trees. *Bioinforma Oxf Engl*. 2009 Aug 1; 25(15):1970–1.
7. Swofford DL. PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). Version 4. Sinauer Associates, Sunderland, Massachusetts. Sunderland, Massachusetts: Sinauer Associates; 2003.
8. Maddison WP, Maddison DR. Mesquite: a modular system for evolutionary analysis. Version 3.2 <http://mequiteproject.org>. 2017.
9. Maddison DR, Schulz K-S. The Tree of Life Web Project. Internet address: <http://tolweb.org>. 2007.
10. Rambaut A. FigTree—Produce Figures of Phylogenetic Trees - <http://tree.bio.ed.ac.uk/software/figtree/>. 2016.
11. Jordan GE, Piel WH. PhyloWidget: web-based visualizations for the tree of life. *Bioinformatics*. 2008 Jul 15; 24(14):1641–2. <https://doi.org/10.1093/bioinformatics/btn235> PMID: [18487241](https://pubmed.ncbi.nlm.nih.gov/18487241/)
12. Rosindell J, Harmon LJ. OneZoom: A Fractal Explorer for the Tree of Life. *PLOS Biol*. 2012 Oct 16; 10(10):e1001406. <https://doi.org/10.1371/journal.pbio.1001406> PMID: [23091419](https://pubmed.ncbi.nlm.nih.gov/23091419/)
13. Felsenstein J. Phylogeny programs. Available at <http://evolution.genetics.washington.edu/phylip/software.html>. 2008.
14. Hughes T, Hyun Y, Liberles DA. Visualising very large phylogenetic trees in three dimensional hyperbolic space. *BMC Bioinformatics*. 2004 Apr 29; 5:48. <https://doi.org/10.1186/1471-2105-5-48> PMID: [15117420](https://pubmed.ncbi.nlm.nih.gov/15117420/)
15. Ware C. Information Visualization: Perception for Design. Elsevier; 2013. 537 p.
16. Meirelles I. Design for Information: An Introduction to the Histories, Theories, and Best Practices Behind Effective Information Visualizations. Rockport Publishers; 2013. 224 p.
17. Johnson B, Shneiderman B. Tree-maps: a space-filling approach to the visualization of hierarchical information structures. In: IEEE Conference on Visualization, 1991 Visualization '91, Proceedings. 1991. p. 284–91.
18. Lü H, Fogarty J. Cascaded Treemaps: Examining the Visibility and Stability of Structure in Treemaps. In: Proceedings of Graphics Interface 2008 [Internet]. Toronto, Ont., Canada, Canada: Canadian Information Processing Society; 2008 [cited 2016 Aug 10]. p. 259–266. (GI '08). Available from: <http://dl.acm.org/citation.cfm?id=1375714.1375758>
19. Hebrard M, Taylor TD. MetaTreeMap: An Alternative Visualization Method for Displaying Metagenomic Phylogenetic Trees. *PLOS ONE*. 2016 Jun 23; 11(6):e0158261. <https://doi.org/10.1371/journal.pone.0158261> PMID: [27336370](https://pubmed.ncbi.nlm.nih.gov/27336370/)
20. Bostock M. D3.js—Data-Driven Documents [Internet]. [cited 2016 Aug 17]. Available from: <https://d3js.org/>
21. Fürbringer M. Bijdragen tot de Dierkunde. Assézat A, editor. Vol. XV. Amsterdam; 1888.
22. Zoomable Circle Packing [Internet]. [cited 2016 Aug 11]. Available from: <http://bl.ocks.org/mbostock/7607535>
23. La Rosa M, Rizzo R, Urso A. Soft Topographic Maps for Clustering and Classifying Bacteria Using Housekeeping Genes. *Adv Artif Neural Syst*. 2011 Oct 12; 2011:e617427.

24. Cortese PF, Battista GD, Moneta A, Patrignani M, Pizzonia M. Topographic Visualization of Prefix Propagation in the Internet. *IEEE Trans Vis Comput Graph*. 2006 Sep; 12(5):725–32. <https://doi.org/10.1109/TVCG.2006.185> PMID: 17080793
25. Davies J. Newick.js. Retrieved from: <https://github.com/jasondavies/newick.js> [Internet]. GitHub. 2011 [cited 2016 Aug 10]. Available from: <https://github.com/jasondavies/newick.js>
26. Right-angle phylograms and circular dendrograms with d3. To preview see <http://bl.ocks.org/kueda/1036776> [Internet]. Gist. [cited 2016 Aug 10]. Available from: <https://gist.github.com/kueda/1036776>
27. Hug LA, Baker BJ, Anantharaman K, Brown CT, Probst AJ, Castelle CJ, et al. A new view of the tree of life. *Nat Microbiol*. 2016 Apr 11; 1(5):16048.
28. Healey CG, Enns JT. Attention and visual memory in visualization and computer graphics. *IEEE Trans Vis Comput Graph*. 2012 Jul; 18(7):1170–88. <https://doi.org/10.1109/TVCG.2011.127> PMID: 21788672