

OPINION ARTICLE

REVISED Animating and exploring phylogenies with fibre plots [version 3; referees: 2 approved, 1 approved with reservations]

William D. Pearse [©]

Department of Biology & Ecology Center, Utah State University, Logan, UT, USA

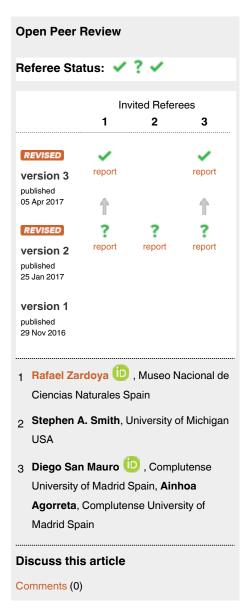
First published: 29 Nov 2016, 5:2790 (doi: 10.12688/f1000research.10274.1)
Second version: 25 Jan 2017, 5:2790 (doi: 10.12688/f1000research.10274.2)
Latest published: 05 Apr 2017, 5:2790 (doi: 10.12688/f1000research.10274.3)

Abstract

Despite the progress that has been made in many other aspects of data visualisation, phylogenies are still represented in much the same way as they first were by Darwin. In this brief essay, I give a short review of what I consider to be some recent major advances, and outline a new kind of phylogenetic visualisation. This new graphic, the fibre plot, uses the metaphor of sections through a tree to describe change in a phylogeny. I suggest it is a useful tool in gaining an rapid overview of the timing and scale of diversification in large phylogenies.



This article is included in the Phylogenetics channel.





Corresponding author: William D. Pearse (will.pearse@usu.edu)

How to cite this article: Pearse WD. Animating and exploring phylogenies with fibre plots [version 3; referees: 2 approved, 1 approved with reservations] F1000Research 2017, 5:2790 (doi: 10.12688/f1000research.10274.3)

Copyright: © 2017 Pearse WD. This is an open access article distributed under the terms of the Creative Commons Attribution Licence, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. Data associated with the article are available under the terms of the Creative Commons Zero "No rights reserved" data waiver (CC0 1.0 Public domain dedication).

Grant information: The author(s) declared that no grants were involved in supporting this work.

Competing interests: The author has no competing interests.

First published: 29 Nov 2016, 5:2790 (doi: 10.12688/f1000research.10274.1)

REVISED Amendments from Version 2

This revision addresses the reviewers' manuscript comments, and incorporates their suggestions into the code that produces a fibre plot (see Supplementary materials). The resulting plot is, I hope, easier to interpret. I am very grateful to all the reviewers for their comments, which have substantially improved this manuscript.

See referee reports

A new generation of phylogeneticists are piecing together the entire tree of life, making vast phylogenies of millions of taxa^{1,2}. Many have produced tree-like depictions of the relationships among species, both before [see 3] and after Darwin described the origin of species⁴, but Haeckel's drawings⁵ are perhaps the most well-known. As our phylogenies become larger, a problem has emerged: humans cannot easily interpret phylogenies with millions of tips. In this brief essay, I will describe recent progress in the visualisation of phylogenies, and outline a new kind of plot—the "fibre plot". My aim is not to write a review [c.f. 6], but rather to provide an opinionated commentary on some major milestones in the progress of phylogenetic visualisation.

Haeckel's phylogenies⁵ are beautiful to look at, and convey the overall structure of a phylogeny well. Each minor branch rarely maps onto a particular species, but their presence reminds the reader of the ever-changing nature of diversification. Both Haeckel and Darwin convey two kinds of information in their visualisations: *time* through depth on the page, and *relatedness* through the branching structure itself. Haeckel is also notable for producing a series of phylogenies, each examining a finer phylogenetic scale. Haeckel grasped that humans cannot process the fine details of all species without becoming lost, and that a series of phylogenies provides the same information in a more digestible format than a single, large, fully-resolved tree.

The last one hundred years have seen transformative changes to phylogenetic inference [see 7], but the same is not true of phylogenetic visualisation. The pace of change of phylogenetic visualisation has not matched that of other aspects of statistical visualisation. A time-traveller from 1859 could decipher a phylogeney from 2017 with *On the Origin of Species*⁴ as a guide, but the box-plots⁸ and histograms⁹ we rely on today would be foreign to them. Circular ("radial") phylogenies are sometimes preferred when space is limited [e.g., 10,11], and "magnifiers" in some computer programs highlight certain parts of the tree in more detail [e.g., 12], but for the most part any advances have been relatively minor.

A major innovation came when programs such as *Walrus*^{13,14} and *Paloverde*¹⁵ allowed users to fly around phylogenies within 3D virtual spaces. Both are notable for presenting structure as something to be *explored*, not merely viewed, and that "a 3D world, offers visual cues that aid in navigation and display that is unavailable in strictly 2D versions of the same layout"¹⁵. The author of *Paloverde*, like Haeckel, recognised that scientists need to shift between finer and coarser phylogenetic scales when examining data, and so

allowed users to collapse nodes at will. These programs were major advances in helping phylogeneticists conceptualise their own phylogenetic hypotheses.

At least as transformative was the release of $OneZoom^{16}$: a fractal phylogeny representation capable (theoretically) of displaying the entire tree of life on one page. OneZoom also requires the user to explore the tree, scanning up and down between finer and coarser details to make sense of the entire tree. Critically, OneZoom's authors recognised that we are reaching the limits of what can be displayed in books: "[w]e now need to take the next step with a transition to data visualization that is optimized for interactive displays rather than printed paper." They suggest that the way to display the next generation of data is to use the next generation of technology.

A common thread running through these developments is their capacity to change the information displayed to the viewer, to better emphasise difference in structure across different phylogenetic depths. Consequently, I suggest the use of a new visualisation, the "fibre plot", which is intended to leverage our natural ability to detect visual change through time. The fibre plot may be considered a horizontal slice through the tree of life, taken at whatever height (depth) the viewer requires (Figure 1). By moving along the tree, from the root to the tip, viewers will see the relative width of each fibre, and so gauge the number of terminal tips subtending that clade. I emphasise that, while Figure 1 shows the underlying logic behind the plot, the "plot" should really be called an animation - it is most readily interpretable when the user watches a video

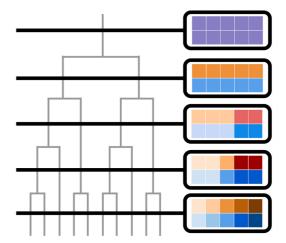


Figure 1. An explanation of a fibre plot. On the left, I show a phylogeny (in grey) with a series of slices cut through it (in black). To the right, I show views through those slices surrounded in black outlines: each of these slices forms the basis of a fibre plot. Within each slice, a square represents descendent tips, and colours of those squares represent the composition of clades within a particular time slice. Squares of the same colour form a "fibre" in the tree of life. A true fibre plot would be an animation of the transition between these slices, showing how the clades (fibres) that make up the tree split as diversification takes place. Alternate colouring schemes are possible for the fibres; the R implementation, by default, colours fibres according to clade age, and allows for different colouring schemes within a plot to highlight taxa of interest.

composed of successive slices through the trunk of the tree. I suggest the animation, with frames recorded at equal intervals along that trunk, provides the viewer with an intuitive sense of the timing of the diversification of major clades. I have written R code to produce a fibre plot (Supplementary File 1; to be released in the package pez¹⁷), and an example of how it can be used to visualise the mammal tree of life¹⁸ (Supplementary File 2). The code can also be used with non-ulatrametric trees, where I find it particularly useful to represent the relative fraction of a tree that is extinct at any given time-point.

Despite humanity being closer than ever to a reliable tree of all life on Earth^{1,2}, phylogenetic visualisation may seem like a niche topic. I strongly feel that phylogenetic visualisation is critical if we are to grasp the full extent of our planet's biodiversity. Human activity has carelessly altered almost every aspect of our planet, and we must now live with the shame and hubris of a geologic age we named after ourselves¹⁹. There has never been a greater need to find a way to show humanity our true place in the world. In whatever sense phylogeneticists have a duty, I believe it is ours to show the world that we are nothing more than a twig on a tree that we are cutting

Author contributions

WDP was responsible for all aspects of this work.

Competing interests

The author has no competing interests.

Grant information

The author declared that no grants were involved in supporting this

Acknowledgements

I am grateful to Bonnie G. Waring for constructive comments and suggestions.

Supplementary files

Supplementary File 1: R code for fibre plot

Click here to access the data.

Supplementary File 2: Mammals.gif. Animated fibre plot in Graphics Interchange Format (GIF) of a phylogeny of all extant mammals.¹⁸ Click here to access the data.

References

- Hedges SB, Marin J, Suleski M, et al.: Tree of life reveals clock-like speciation and diversification. Mol Biol Evol. 2015; 32(4): 835-845. PubMed Abstract | Publisher Full Text | Free Full Text
- Hinchliff CE, Smith SA, Allman JF, et al.: Synthesis of phylogeny and taxonomy into a comprehensive tree of life. Proc Natl Acad Sci U S A. 2015; 112(41): PubMed Abstract | Publisher Full Text | Free Full Text
 - Pietsch TW: Trees of life: a visual history of evolution. JHU Press; 2012.
- Darwin C: On the origin of species. John Murray, London; 1859.
- Haeckel EH: Generelle Morphologie der Organismen allgemeine Grundzuge der organischen Formen-Wissenschaft, mechanisch begrundet durch die von Charles Darwin reformirte Descendenz-Theorie von Ernst Haeckel: Allgemeine Entwickelungsgeschichte der Organismen kritische Grundzuge der mechanischen Wissenschaft von den entstehenden Formen der Organismen. begrundet durch die Descendenz-Theorie. Verlag von Georg Reimer, 1866; 2.
- Page RD: Space, time, form: viewing the Tree of Life. Trends Ecol Evol. 2012; 27(2): 113-120. PubMed Abstract | Publisher Full Text
- Felsenstein J: Inferring phylogenies. Sinauer Associates; 2004. Reference Source
- Tukey JW: Exploratory data analysis. Reading, Mass; 1977.
- Ioannidis Y: The history of histograms (abridged). Proceedings of the 29th international conference on Very large data bases-Volume 29. VLDB Endowment, 2003: 19-30 Publisher Full Text

- Paradis E, Claude J, Strimmer K: APE: Analyses of Phylogenetics and Evolution in R language. Bioinformatics. 2004; 20(2): 289-290 PubMed Abstract | Publisher Full Text
- 11. Rambaut A: Figtree. 2017.
- Huson DH, Scornavacca C: Dendroscope 3: an interactive tool for rooted phylogenetic trees and networks. Syst Biol. 2012; 61(6): 1061-1067. PubMed Abstract | Publisher Full Text
- Munzner T: H3: Laying out large directed graphs in 3d hyperbolic space. In 13. Information Visualization, 1997. Proceedings, IEEE Symposium. IEEE, 1997; **Publisher Full Text**
- Hughes T, Hyun Y, Liberles DA: Visualising very large phylogenetic trees in three dimensional hyperbolic space. BMC Bioinformatics. 2004; 5(1): 48. PubMed Abstract | Publisher Full Text | Free Full Text
- Sanderson MJ: Paloverde: an OpenGL 3D phylogeny browser. Bioinformatics. 2006; 22(8): 1004-1006. PubMed Abstract | Publisher Full Text
- Rosindell J, Harmon LJ: OneZoom: a fractal explorer for the tree of life. PLoS Biol. 2012; 10(10): e1001406. PubMed Abstract | Publisher Full Text | Free Full Text
- Pearse WD, Cadotte MW, Cavender-Bares J, et al.: pez: phylogenetics for the environmental sciences. Bioinformatics. 2015; 31(17): 2888-2890. PubMed Abstract | Publisher Full Text
- Bininda-Emonds OR, Cardillo M, Jones KE, et al.: The delayed rise of present-day mammals, Nature, 2007; 446(7135); 507-12. PubMed Abstract | Publisher Full Text
- Crutzen PJ: Geology of mankind. Nature. 2002; 415(6867): 23. PubMed Abstract | Publisher Full Text

Open Peer Review

Current Referee Status:







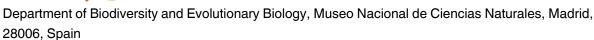
Version 3

Referee Report 19 April 2017

doi:10.5256/f1000research.12226.r21588



Rafael Zardoya 🗓



The revised version of this article looks clearly improved as well as the new output of the fibre plot, which is now easier to interpret. Now, users have the last word on how popular and useful this software may become.

Competing Interests: No competing interests were disclosed.

Referee Expertise: Phylogenetics, comparative genomics, zoology

I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Referee Report 10 April 2017

doi:10.5256/f1000research.12226.r21586



Diego San Mauro 🗓 , Ainhoa Agorreta

Department of Zoology and Physical Anthropology, Complutense University of Madrid, Madrid, Spain

In our first review of this opinion article, we indicated several comments and suggestions for a few aspects that required further attention. We see that the author has incorporated most of the changes suggested (either by us or the other two referees) into the article, and we feel it looks now improved. In our opinion, this new version of the paper is satisfactory, and suitable for publication in F1000Research.

Competing Interests: No competing interests were disclosed.

We have read this submission. We believe that we have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Version 2



Referee Report 17 February 2017

doi:10.5256/f1000research.11550.r19938

Diego San Mauro (D), Ainhoa Agorreta

Department of Zoology and Physical Anthropology, Complutense University of Madrid, Madrid, Spain

This paper presents a brief overview on phylogenetic visualization and introduces a novel approach for visualizing phylogenies (timetrees) using fibre plots. Given the rapid accumulation of phylogenetic information over the last years that has enabled the construction of massive trees (mega-phylogenies) containing millions of branches and leaves (taxa), the new visualization method appears to be interesting and with some potential. However, it is outlined only succinctly in the paper, and we feel that there are a few issues that require further attention. More discussion is needed on the specific applications and/on implementations of phylogenetic fibre plots compared to the other visualization approaches already available. For instance, what are the advantages of fiber plots over conventional phylogenetic plots in terms of comparing e.g. different topology sets? (as used for example in hypothesis testing). Also, what is the applicability (if any) of fiber plots for visualizing phylogenetic trees whose branches represent rate of evolution (e.g., substitutions/site) instead of time? (as in phylograms). Or, how do fibre plots deal with extinct branches? (as those displayed by extinct fossil lineages). Discussing these issues (among others) more in detail would make it easier for the reader to assess the breadth of novelty and usefulness of the new method for the general field of phylogenetics, and its applicability beyond the reconstruction of the timetree of life. As described in the current paper, it seems that fibre plots could be a complement, but not substitute of the other (more conventional) phylogenetic visualization approaches. The output of the fibre plot is colorful, but in general very difficult to interpret. In fact, interpreting the fibre plot output of very large phylogenies or even the tree of all life would be more difficult than interpreting more conventional approaches (those zooming in and out the phylogeny). Implementing some sort of labeling/cross-referencing with lists of taxa or even conventional phylogenetic trees live on the side could help in the precise interpretation of what is being displayed at each timeframe.

There are also some additional issues that we want to mention:

- First paragraph: The sentence beginning "Many have..." needs some rewording... It is true that many have produced tree-like depictions of the relationships among species, but certainly not many before Darwin. So, please reword.
- Second paragraph: Please provide a reference for the sentence beginning "Haeckel grasped that humans...".
- Third paragraph: Besides Dendroscope, it would be fair to cite FigTree (http://tree.bio.ed.ac.uk/software/figtree/) as well in the last sentence.
- Fourth paragraph: It would be good to cite and discuss Walrus (http://www.caida.org/tools/visualization/walrus/) here as well. It appeared in 2001 (earlier than Paloverde) and allowed interactive 3D visualization of hierarchical graphs.
- Fifth paragraph: Please add references and expand the last statement about using Hilbert curves.
- Last paragraph: The last paragraph of the paper appears unnecessary and probably should be removed. Only the first sentence could be kept as part of the previous paragraph (as closing



statement). If this sentence is retained, please keep in mind that phylogenies (e.g., the tree of life) are hypotheses. Therefore, it would be more appropriate to say "...being closer than ever to a reliable tree of all life", rather than "...being closer than ever to a true tree of all life".

Competing Interests: No competing interests were disclosed.

We have read this submission. We believe that we have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however we have significant reservations, as outlined above.

Author Response 29 Mar 2017

Will Pearse, Utah State University, USA

I thank you both for your comments, which have greatly improved the article. I'm particularly grateful that you mentioned Walrus; this was a huge oversight on my part, and I'm glad to have an opportunity to correct it! I apologise for that, due to space limitations of an opinion article (limited to 1000 words), I am not able to go into as much detail as I would like on some of the broader topics you raise. I have, however, significantly altered the code of the fibre plot following your suggestion about non-ultrametric phylogenies and highlighting particular taxa. In particular, your suggest of a phylogeny to the side of the plot, mirroring reviewer 2's suggestion, has greatly improved the figure. Thank you!

Responding to each of your comments in turn:

- Branch lengths and extinct taxa. I have re-written the function so that it supports dated and undated trees, and highlights extinct taxa to show the time period within which they went extinct. I describe this in the penultimate paragraph of the manuscript.
- Ease of interpretation and suggestion of replacement of other phylogenies. I agree with the reviewers that this is not a replacement for a traditional visualisation; as I discuss in the text I find the visualisation captures well changes in timing and diversification more readily in extremely large phylogenies (e.g., the ~5000 taxon example I provide). I have followed the reviewers' suggestions and allowed the user to highlight clades and taxa of interest, which, along with the comments of reviewer 3, I hope make the plot easier to interpret.
- First paragraph: I respectfully disagree with your comment; the book by Pietsch I reference contains many examples of tree-like structures preceding Darwin. I have altered the text to make my meaning clearer.
- Second paragraph: Thank you for this; I have added a reference earlier in the paragraph.
- Third paragraph: Thank you for this; I now cite FigTree and the R package ape.
- Fourth paragrph: Thank you for this; I now mention Walrus (citing a 1997 conference paper that describes what is essentially the same software under the name 'H3'), and cite another software package that converts phylogenies into Walrus format.



- Fifth paragraph: Thank you for this; having now experimented more thoroughly with the approach, I didn't find it aided interpretation. I have changed the code to alter the layout of the fibres, but I have dropped this reference from the text.
- Final paragraph: Thank you for this; I have made the changes your suggested.

Competing Interests: No competing interests were disclosed.

Referee Report 13 February 2017

doi:10.5256/f1000research.11550.r19876

Stephen A. Smith

Department of Ecology and Evolutionary Biology, University of Michigan, Ann Arbor, MI, USA

Pearse presents a new means for visualizing large phylogenies called the fibre plot. The purpose of this plot is to better represent splits by different colors and shades. This is an interesting idea and is demonstrated with an accompanying animated gif. However, I am left wondering if there are significant insights gathered from this view of the mammal tree. The animation proceeds and areas of the graph change color. I understand why they change but don't know where I am in the tree and what the significance of the change is in speed or area of the graph. The figure presented (Figure 1) shows a somewhat different view of the fibre plot as presented alongside the phylogeny. This makes me think that perhaps a more informative presentation would be the view of the phylogeny along with the fibre plot. Then the animation would follow a line that moves in a preorder fashion from the root to the tips. This would allow for a more direct comparison of the tree and the plot. Without this additional guide, I am not sure what to make of the animation. I don't know where I am in the tree (in time or place) and I can't "move around" in any particular way. I can also envision any number of statistics presented with the plot.

This is an interesting start of an idea but I think it needs a little more development before it would be useful for navigating the size of the tree intended by the author. However, there may be some interesting uses for this or something like it in the future.

Editorial comments

I recommend that the author edit the abstract. For example, the sentence "Despite the progress that has been made in the visualisation of information since Haeckel's time, phylogenetic visualisation has moved forward remarkably little." seems to suggest that Haeckel was the first person to try and visualize data. While this may be accurate for some biological data, it is not true for data in general as cartographers have been trying to visualize information and data for centuries. The final sentence in the paragraph could also use some adjustments. While the statement is trying to convey a general sense of the importance of phylogenies, I am not certain that "our place" in the tree of life will dramatically change as a result of visualization of the data. I would also recommend changes to the intervening sentences.

I would recommend some changes to the remaining text but won't outline all of those here.

Competing Interests: No competing interests were disclosed.



I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Author Response 29 Mar 2017

Will Pearse, Utah State University, USA

Thank you for these suggestions. I have now altered the fibre plot code in exactly the way you suggested, adding a traditional phylogeny to the right-hand side of the animation that grows and matches colour with the fibre plot. I think it makes the plot much easier to interpret - thank you for this excellent suggestion! I have edited the abstract following your suggestions, and removed the final sentence from it entirely.

Competing Interests: No competing interests were disclosed.

Referee Report 10 February 2017

doi:10.5256/f1000research.11550.r19939

Rafael Zardoya (10)



Department of Biodiversity and Evolutionary Biology, Museo Nacional de Ciencias Naturales, Madrid, 28006, Spain

The paper entitled "Animating and exploring phylogenies with fibre plots" by Pearse is an interesting contribution that proposes a new and distinct way to visualize phylogenetic trees. The new method propose by the author uses fibre plots to slice a phylogenetic tree from root to tips and visualize, as an animation, the cladogenetic process in time.

As the author correctly argues, while it is now possible to reconstruct phylogenetic trees involving tens of thousands of species, visualization of such trees is complex and has not advanced at the same pace as probabilistic inference methodology. Hence, the challenge is set. There are many programs for visualizing trees but few have explored the need of dealing with large phylogenies. Different strategies have been proposed to represent phylogenies including the collapse of certain nodes, distortion of the view, and representation in 3D, but thus far, the most popular approach probably consists on zooming in and out the phylogeny (OneZoom, Rosindell et al. 2012) using appropriate tools (e.g., a tablet). These viewers are complemented with others that allow incorporating other information pertinent to the phylogeny (e.g., iTOL, Letunic and Bork 2016).

The proposal here presented explores in a very different direction. While the idea of looking at different temporal slices in the phylogeny to get a feeling of the timing of diversification of the different clades is original, I think it is too preliminary in the present contribution. The video composed of successive slices shows in different colors how a single (ancestral) lineage is successively split into many but the viewer is unable to discern to which exact descendant lineages is looking at, as there are no labels. Moreover, at some point the number of splits (and colors) is too large to obtain useful information from the animation. As presently devised, the analysis of different clades will render very similar plots, which will be difficult to interpret (beyond seen an increase in the number of lineages) and compare. If the author wants this tool to be widely used, he should make the final outcome more appealing and understandable (e.g., perhaps a



grid plot with labels of each lineage in the corresponding axis would help following which lineages and their ancestors are diverging) by peers from other fields than phylogenetics and by the general public.

Minor changes:

- The author mentions that PaloVerde in 2006 was the first 3D phylogeny viewer to his knowledge. I think he should check the Walrus graph visualisation tool by Hughes et al. 2004
- Close brackets after [see 6]

References

- 1. Rosindell J, Harmon LJ: OneZoom: a fractal explorer for the tree of life. *PLoS Biol.* 2012; **10** (10): e1001406 PubMed Abstract I Publisher Full Text
- 2. Letunic I, Bork P: Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Res.* 2016; **44** (W1): W242-5 PubMed Abstract I Publisher Full Text
- 3. Hughes T, Hyun Y, Liberles DA: Visualising very large phylogenetic trees in three dimensional hyperbolic space. *BMC Bioinformatics*. 2004; **5**: 48 PubMed Abstract | Publisher Full Text

Competing Interests: No competing interests were disclosed.

I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Author Response 29 Mar 2017

Will Pearse, Utah State University, USA

Thank you for these comments. I hope the changes I have made address some of your concerns about interpretation, all of which I think are legitimate. I have added the capacity to label (and track) particular species and clades through the animation, and have added an optional display of the phylogeny to the side of the animation. I hope the reviewer agrees that this addresses their concerns. Your comments about the colouring scheme, in particular, were very useful - I now colour everything according to clade age, which I hope you will agree makes for a much more informative plot. Thank you also for mentioning Walrus - omitting this was a huge mistake, and I'm grateful you've corrected me.

Competing Interests: No competing interests were disclosed.