

Visualization of evolutionary relationships through geologic time in *Timescale Creator*



Abdullah Khan Zehady^{a,1}, James G. Ogg^{b,a,*,2}, Barry G. Fordham^{c,3}, Gangi Palem^{d,4}, Jason Bobick^{e,5}, Gabi M. Ogg^{f,6}

^a Department of Earth, Atmospheric and Planetary Science, Purdue University, West Lafayette, IN, 47907, USA

^b State Key Laboratory of Oil and Gas Reservoir Geology and Exploitation, Chengdu University of Technology, Chengdu, Sichuan, 610059, China

^c Research School of Earth Sciences, Australian National University, Canberra, ACT, 2601, Australia

^d Google Inc., 1600 Amphitheatre Parkway, Mountain View, CA, 94043, USA

^e Department of Computer and Information Technology, Purdue University, West Lafayette, IN, 47907, USA

^f Geologic TimeScale Foundation, 1224 N. Salisbury St., West Lafayette, IN, 47906, USA

ARTICLE INFO

Keywords:

Paleontology
Fossil record
Education
Computer graphics
Data visualization

ABSTRACT

Evolutionary trees showing interpretations of the divergences, lineages, extinctions and relative abundances of life forms through geologic time provide a very useful visualization for both public audiences and research communities. The Java-based *TimeScale Creator* platform was originally designed to present graphics of Earth history with a streamlined interface. We have added an evolutionary tree display column in addition to the array of other geo-history columns for chart making and on-screen exploration through any interval of geologic time. Features in the evolutionary tree display include on-screen images of life forms, single-click on/off control of color-coded family branches, pop-up windows giving additional details, embedded evolution charts of morphotypes that developed within genotypes, an option to make more-compact charts by allowing new evolutionary branches to reposition over earlier extinct branches, and export/import of some common evolution-data sharing formats. A suite of current public datapacks using this evolutionary-tree visualization include major vertebrate groups (including dinosaurs), marine plankton through the past ca. 70 million years, and the Tree of Life of extant organisms. Students and researchers can upload their own datasets using a simple tabular-data format from Excel-type spreadsheets. The package and external datapacks are freely available at www.timescalecreator.org. The intended audiences range from Earth-science classes in schools to paleontologists and evolutionary biologists who wish to compare their phylogenies to other geologic events and trends of Earth history.

1. Introduction

The developments of the diversity of life on Earth are routinely illustrated by evolutionary trees. Even though an evolutionary tree is an interpretation based on biological or paleontological evidence subject to

continual revision, it does provide a vivid diagram and a model that can be correlated to other events in Earth history in order to gain insights on environmental and other feedbacks such as global carbonate trends, major impacts, tectonics, volcanism etc.

The *TimeScale Creator* visualization software displays from an

* Corresponding author. State Key Laboratory of Oil and Gas Reservoir Geology and Exploitation, Chengdu University of Technology, Chengdu, Sichuan, 610059, China.

E-mail addresses: azehady@cisco.com (A.K. Zehady), jogg@purdue.edu (J.G. Ogg), barry.fordham@anu.edu.au (B.G. Fordham), palemgangireddy@gmail.com (G. Palem), jbobick@purdue.edu (J. Bobick), gabiogg@hotmail.com (G.M. Ogg).

¹ Programmed the evolutionary tree software and authored the drafts of the manuscript.

² Supervises the *TSCreator* program and co-wrote the final manuscript.

³ Guided the development of evolutionary tree visualization within the *TSCreator* system and compiled the planktonic foraminifer evolution database for this project.

⁴ Developed the initial version of evolution display.

⁵ Technical assistant for *TimeScale Creator* software project.

⁶ Coordinated testing and user-interfaces for the display of evolutionary trees, prepared the final versions for the *TSCreator* website and enhanced the graphics.

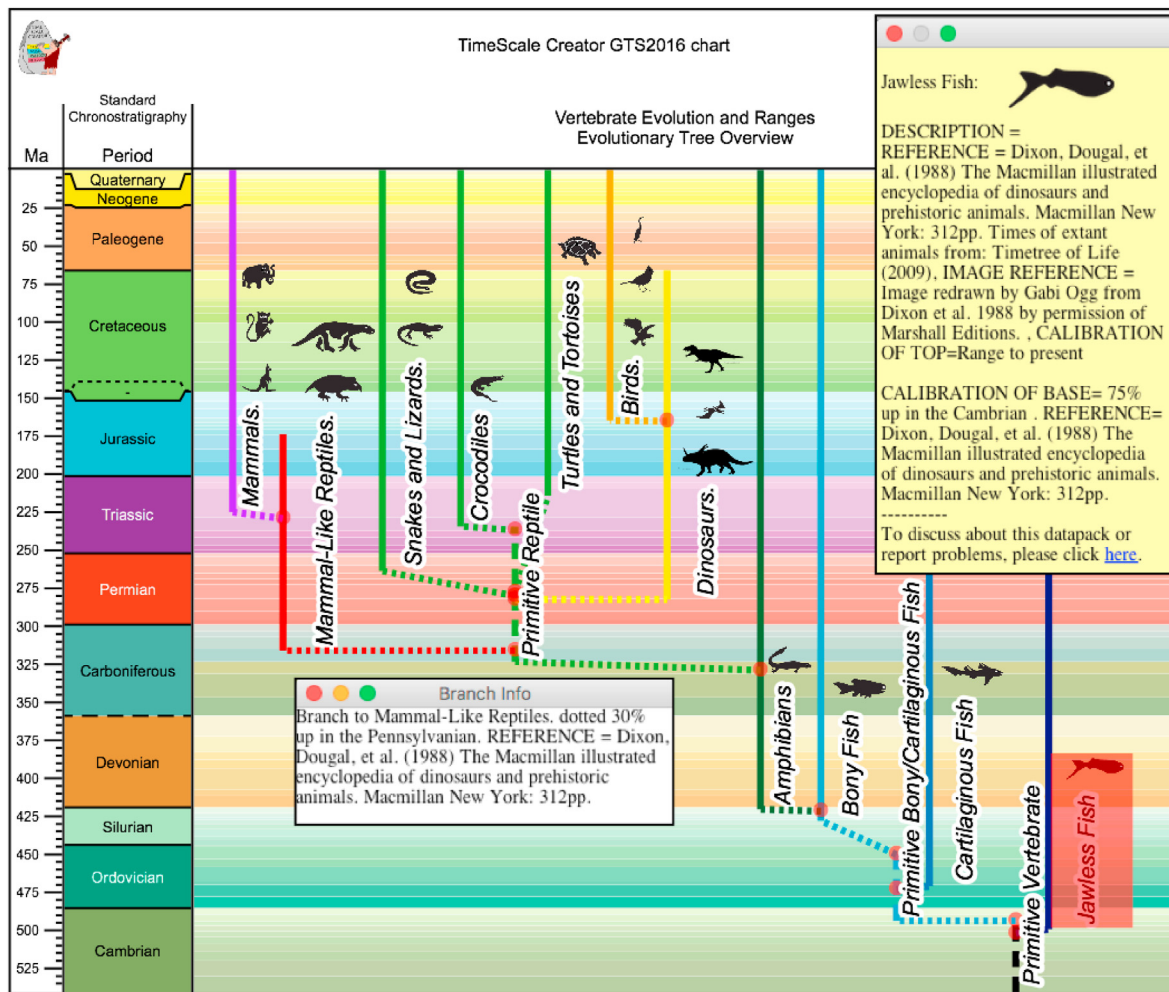


Fig. 1. Example of evolutionary tree display within the TimeScale Creator system using a subset from a generalized vertebrate datapack (fish, dinosaurs, birds, mammals, etc.). The datapack contains color-coded ancestor–descendant relationships and ranges, a library of images, and details on each group. A left-click on the activated red-shaded highlight (e.g., Jawless Fish in this example) opens a pop-up window (yellow rectangle) containing details, images and URL links. Right-clicking on an evolutionary node (red dots on ranges) opens a pop-up with details about that ancestor–descendant relationship. This chart has used the option to display a background of pastel colors (international standard; Commission for the Geological Map of the World, 2012) for the stages within each geologic period.

extensive suite of detailed Earth-history information (regional stratigraphy, geomagnetic polarity, marine macro- and micro-fossil evolution, vertebrate and hominid evolution, stable isotopic and sea-level curves, global reconstruction images, etc.) that have been migrated to standardized geologic time scales for comparative analysis. We were motivated to design a new component that would incorporate an automated production of phylogenetic trees for ancient fossil branches and extant life forms that would provide other features requested by users, such as images, pop-up windows displaying additional information on organisms and details of the age calibrations, and simplification of large trees by collapsing subtrees and growing new branches over extinct ones (Fig. 1). This evolutionary tree visualization suite is for both educational and research audiences.

1.1. Selected related phylogenetic software

The dynamic field of phylogenetic research had led to a huge and growing array of software tools (e.g., one website posting in 2014 gave annotated descriptions of 392 phylogeny packages and 54 free web-based systems (Felsenstein, 2014). We will only highlight a few selected or recent software tools to compare methodologies.

Tree reconstruction is performed by mathematical modeling based on the evolutionary data in some software packages, and model selection is

necessary to identify the best model that can generate the most accurate tree. Some of the leading software package, such as BayesTraits (Meade and Pagel, 2019), MrBayes (Huelsenbeck and Ronquist, 2001), BEAST (Drummond and Rambaut, 2007), and TOPALi (Milne et al., 2009), take a probabilistic approach by using Bayesian inference methods. Others, such as PHYLIP (Felsenstein, 2019), phyclust (Chen, 2019), MetaPIGA2 (Helaers et al., 2010) and PhyML (Guindon et al., 2010) include tools for estimating the likelihood of generated trees using maximum likelihood or use AIC (Akaike Information Criterion) or BIC (Bayesian Information Criterion) for model selection. Tree reconstruction procedures using distance matrices are used in other software, such as BioNumerics (Applied Maths, 2020) and phangorn (Schliep, 2010). However, the aforementioned software packages lack the capability to display evolutionary trees using geologic time. Mesquite (Maddison et al., 2018) is one of the few software packages besides TimeScale Creator that can be used to draw an evolutionary tree with an adjacent geologic timescale, but it is designed to emphasize phylogenetic analysis rather than a detailed visualization of lineages of organisms.

The software packages for tree visualization are written in a variety of programming languages and libraries. Packages written in the R programming language to reconstruct, visualize and perform statistical analysis with the tree data include Rphylip (Revel and Chamberlain, 2014), GGTREE (Yu, 2017), phyttools (Revell, 2012) and Phybase (Liu

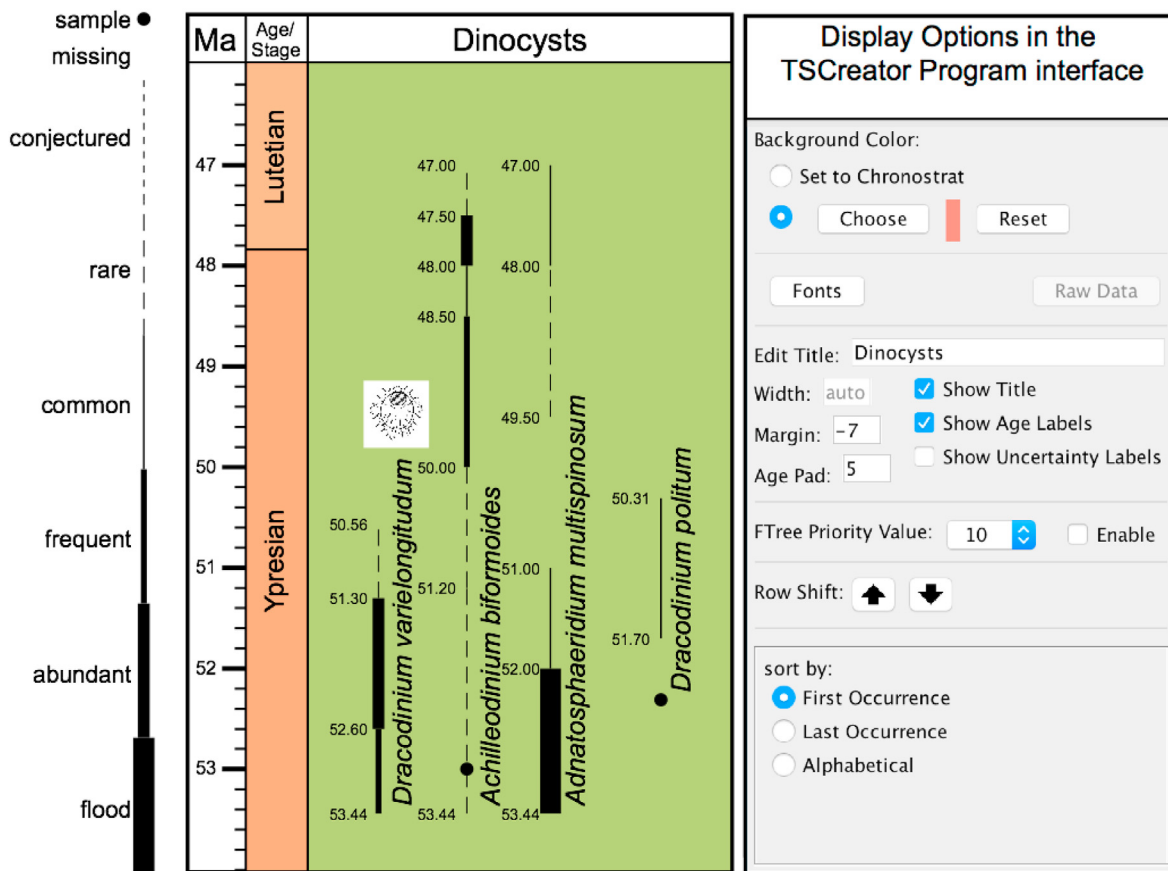


Fig. 2. Stratigraphic range chart in TSCreator using a schematic set of dinoflagellates from 54 to 46 million years ago (Ma). A portion of the underlying dataset array is given in Appendix A. The program auto-adjusts the horizontal spacing between ranges if their labels include images. Options in the menu on the right side include background color, font size and color, editing titles, age labels (activated here), positions of range labels, displaying only those ranges that exceed an assigned priority (useful when a chart has hundreds of ranges), and arrangement of the ranges by first/last occurrence or alphabetically.

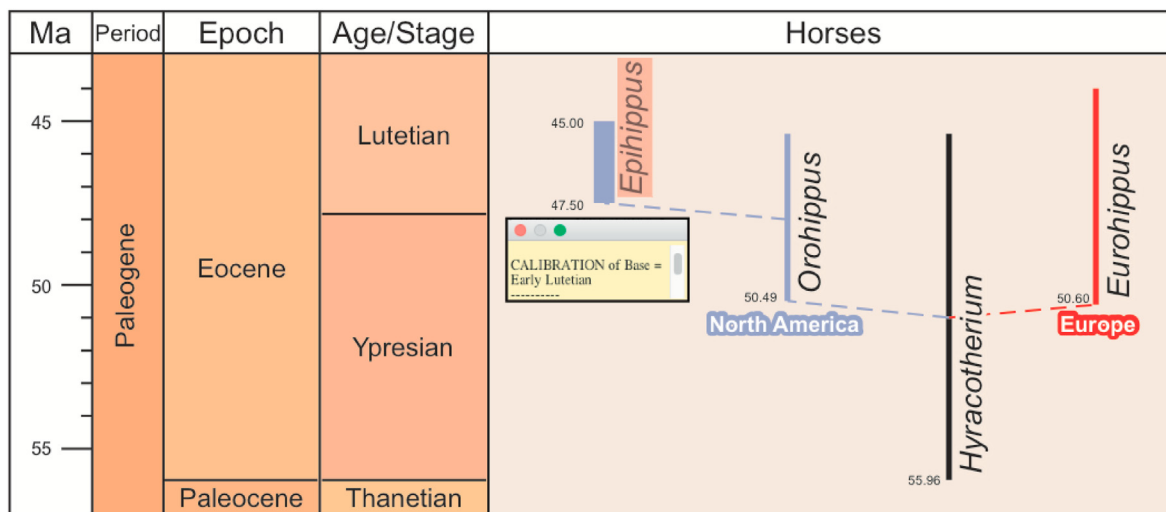


Fig. 3. A portion of a schematic evolution of horses generated from the sample datapack of Appendix B. In the display, a mouse hovering over the “Epihippus” range activated the orange background and then was clicked to open the yellow-shaded pop-up window with information on age calibration (item in column E in the dataset in Fig. A2 example). Connecting branch lines have been set as “dashed”. Subtrees labels (North America, Europe) have colors which are also inherited by the component ranges of that subtree. The display of the ages for range bases has been turned on (see Fig. 2).

and Yu, 2010). Some Javascript libraries, such as phylotree.js (Shank et al., 2018), jsPhyloSVG (Smits and Ouverney, 2010), and PhyD3 (Kreft, 2017), are developed to create websites for phylogenetic tree visualization of publishable quality. Online tools such as AQUAPONY (Cazaux,

2019), EvolView (Zhang et al., 2012), IcyTree (Vaughan, 2017), iTOL-interactive Tree Of Life (Letunic and Bork, 2006), Phylo.io (Robinson et al., 2016) and T-REX (Alix et al., 2012) can be accessed via modern web browsers, and user-formatted tree-files can be uploaded for

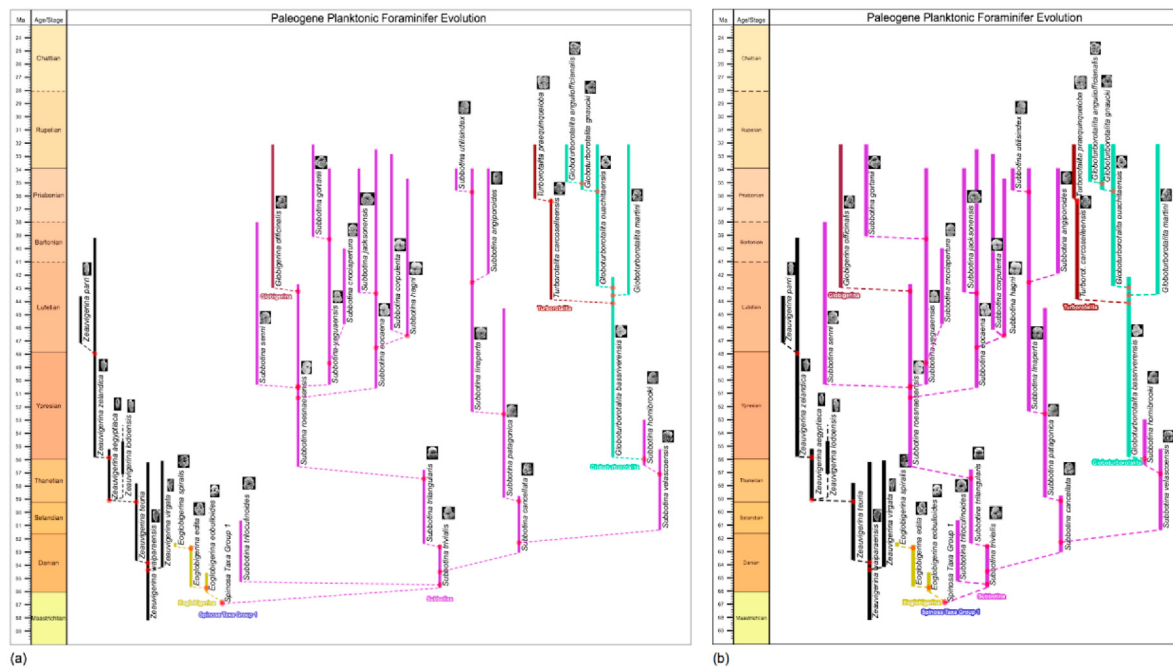


Fig. 4. Conserve chart space option. (a) An extract from the Paleogene Planktonic Foraminifer datapack without the space conservation feature. There is empty space above earlier branches due to the extinction of some families, such as the *Eoglobigerina* (yellow range lines in the Danian Stage). (b) Activating the “Conserve Chart Space” option reduces the needed display space. Some *Subbotina* clades (pink range lines) are now shifted above the extinct *Eoglobigerina* set to enable a more compact chart.

visualization in some of these on-line systems. Examples of some desktop software packages which provide integrated environments for interactive tree visualization and annotation for different operating systems are UGENE (Okonechnikov, 2012), PHYLOVIZ (Francisco et al., 2012), TreeGraph2 (Stöver and Müller, 2010) and Dendroscope (Huson et al., 2007). There has been a trend to partly standardize the uploading of tree data for use in these varied systems using formats called Newick, Nexus, FASTA, etc.

Many of these software packages are dedicated for specific purposes with strengths in particular tasks or static visualization (e.g. compatibility analysis, tree-based sequence alignment, biogeographic analysis, tree simulation, etc.). In contrast, the new tree feature developed for the *TSCreator* software system focuses on deep-time tree visualization alongside extensive datasets on Earth history.

1.2. TimeScale Creator platform for Earth history

TimeScale Creator (TSCreator) is a free Java package for depicting Earth history and deriving custom charts of any portion of the geologic time scale. The visualization interfaces and extensive datapacks were developed and progressively enhanced in conjunction with stratigraphers and paleontologists, including the International Commission on Stratigraphy and several geological surveys (e.g., Gradstein and Ogg, 2006; Ogg and Przybylski, 2009; Ogg et al., 2010; Smith et al., 2015; Bobick et al., 2016). The *TSCreator* internal database suite now contains over 300 columns with over 50,000 events or data points on sea level, stable isotopes, paleobiology, geomagnetics and other aspects of Earth history (plus the Moon and Mars). In addition to this internal dataset, users can incorporate any of about twenty specialized datapacks (e.g., biostratigraphic and geologic history of all Australian basins as compiled by Geoscience Australia (Smith et al., 2015)). In addition to interactive screen displays, the *TSCreator* can export charts in different formats.

Pre-2008 versions had focused on a convenient user-friendly way to generate charts. A series of *TSCreator* updates progressively added features, such as: pop-up windows with images; map interfaces to access regional stratigraphic columns and transects; an age-slider option to

display regional rock types in relation to time; depth-to-age conversions; search tools; and web-accessed datapack makers. All datapacks are based on tab-delimited text and can be prepared in Microsoft Excel, which also provides a flexible method to automatically recalculate the ages of all events to new published age models. Many data-column types (e.g., zones, first/last occurrence arrows, geochemical curves, lithology patterns, etc.) can be combined in a single text file separated by headers to instruct the visualization software. *TSCreator* version 7 and the extensive array of external datapacks are currently standardized to the published age models of *Concise Geologic Time Scale 2016* (Ogg et al., 2016), and are now being recalibrated to *The Geologic Time Scale 2020* (Gradstein et al., 2020).

2. Methods

2.1. Implementation of the tree visualization in TimeScale Creator

The earlier *TSCreator* versions had the capability to display the stratigraphic ranges of organisms against the geologic time scale but lacked a means to show postulated evolutionary relationships among those organisms. The basic range diagram is a set of labeled vertical lines that begin and end at designated ages and have line styles denoting relative abundances (Fig. 2; see Appendix A for this datapack format).

The basic concept in an evolutionary tree is that a “Parent” (ancestor) produces a “Child” (descendant). Parents can have multiple Childs. The initial approach to build an evolutionary tree within *TSCreator* was merely to add a “branch” option to the existing range display (Fig. 2) as a flag to connect a point along the Parent range to the base of the named Child range. Once implemented, other display enhancements were made in response to user recommendations such as more information in the popup boxes, age labels.

We wished to be able to employ alternative tree styles to represent a variety of approaches to phylogeny reconstruction. These include trees of relationships, such as cladograms, and trees with time-relationships with an implied chronology. We also wished the flexibility to display either species-level taxa or higher taxonomic groups to understand evolutionary

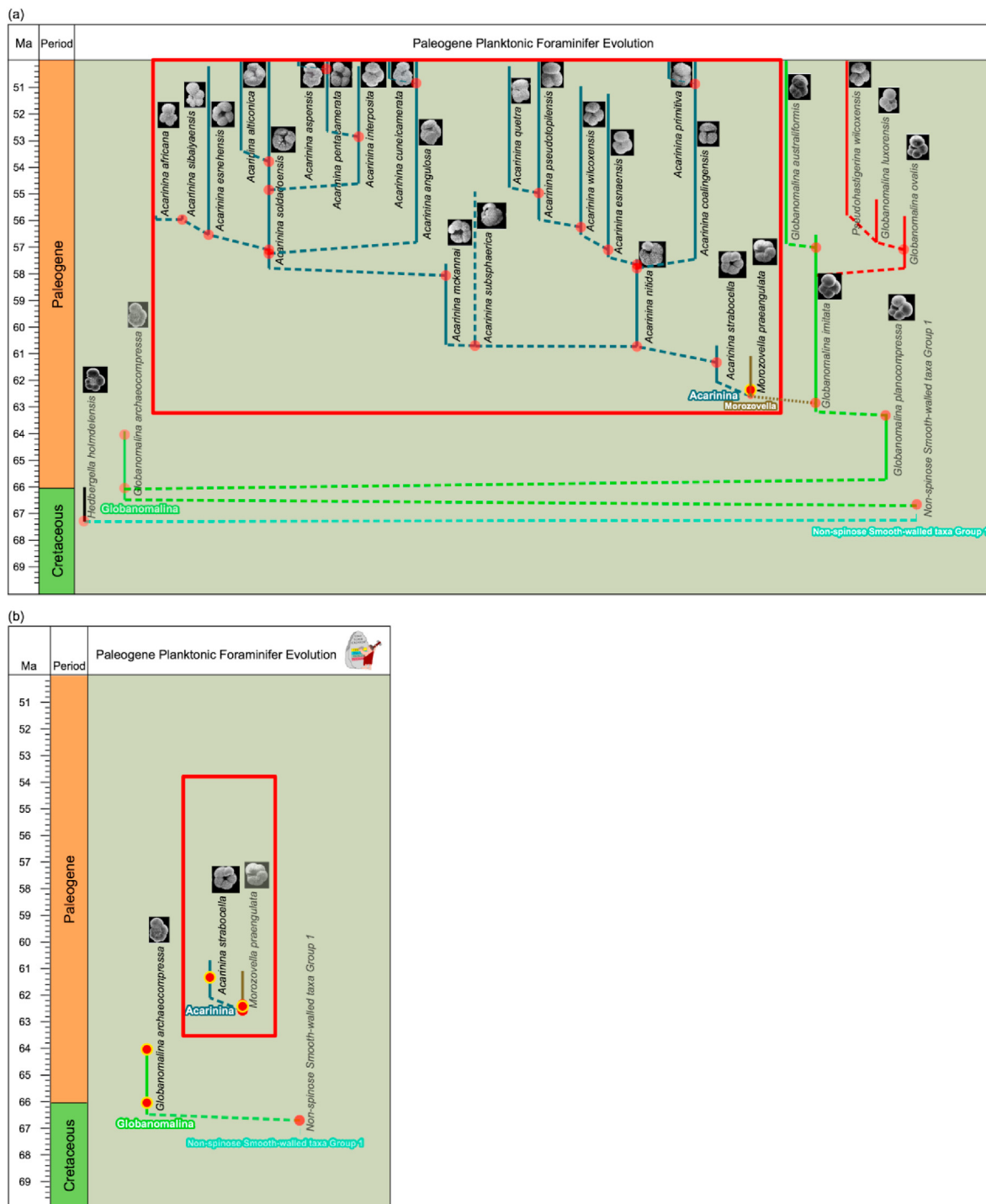


Fig. 5. Using “Priority” assignments in a complex evolutionary tree. (a) A portion of the evolutionary tree of Paleogene Planktonic Foraminifera focusing on the *Acarinina* clade (within red rectangle along with ancestor, *M. praeangulata*; the two *Morozovella* clades descending from *M. praeangulata* are turned off, indicated by hidden nodes with yellow borders). This was generated without enabling “FTree Priority Value”. (b) The same portion of the evolutionary tree, in which FTree Priority Value is enabled and the value has been set to 10, thereby displaying only subtrees with priority values of 10. The display of the *Acarinina* clade is only its initial *A. strabocella* because the other *Acarinina* morphospecies have been assigned lower default priorities in the database and so are now represented by hidden nodes (with yellow borders).

macrotrends.

2.1.1. Data structure for evolutionary tree

The addition of a “branch” flag is all that was required to augment the pre-existing coding of range data in *TSCreator* to add interconnections to produce an evolutionary tree display. For example, to show the evolution of Birds from Dinosaurs, a branch line in the Late Jurassic (160 Ma)

connects that age-point on the Dinosaur range (indicated by a “red-dot”) to the base of the earliest fossil record of Birds at 150 Ma (Fig. 1). This branch flag was given additional enhancing options including: branch labels with horizontal-oriented text for subtrees (clades) or other groupings (e.g., “North America” in Fig. 4), assigning colors to these labels and associated subtree range lines, on/off default settings for subtrees to allow initial pruning when displaying very large charts, branch

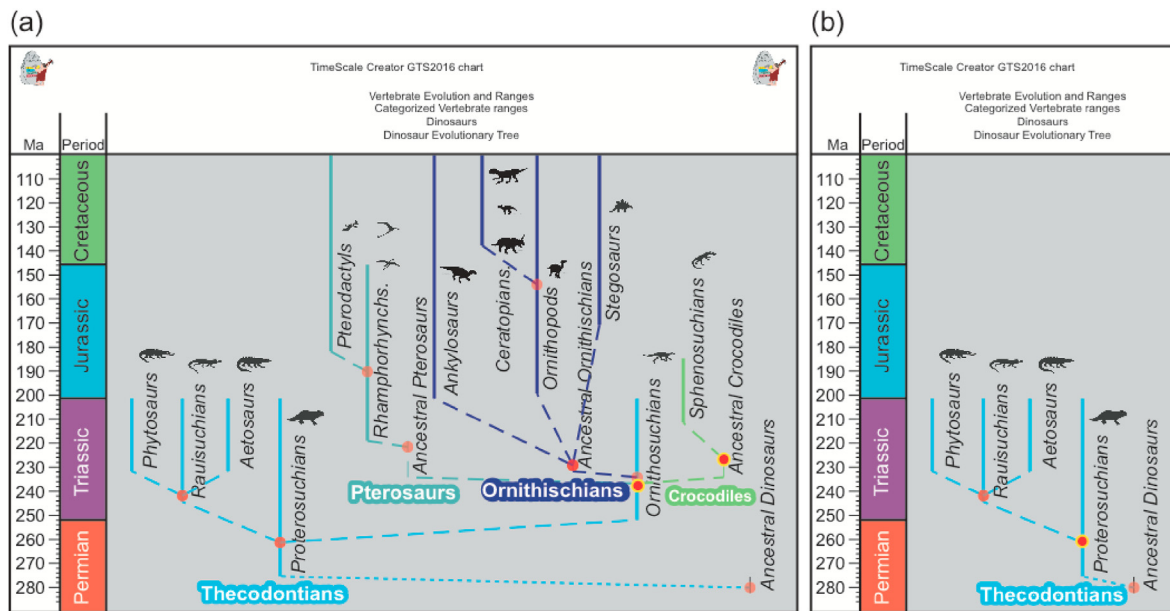


Fig. 6. Turning branches on/off. (a) Schematic evolution of early land reptiles and their dinosaur and crocodile descendants. The “red-dot” nodes (without yellow borders) on range lines at the beginning of evolutionary branches can be left-clicked on/off. (b) The dinosaur branch (Pterosaurs, Ornithischians) and the crocodile branch are not displayed after left-clicking the red-dot on the Proterosuchians range; and the red-dot continues blinking (indicated here by a yellow-border to the red dot) to indicate a turned-off subtree can be activated by left-clicking on that blinking red dot.

line styles (solid, dashed, dotted), and pop-up window text (plus images, URLs) attachable to any taxon (range line) to provide on-screen information (e.g., Figs. 1 and 3, and Fig. A2 in Appendix B).

2.1.2. Algorithm and programming the tree display

The display of evolutionary trees required several steps.

- **Input:** Evolutionary tree datapack in text format.

- **Output:** On-screen SVG display with the evolutionary tree column, and SVG or PDF export.

- **Steps:**

- (1) **Data extraction.** Extract evolutionary-range and branch point data according to the tree data structure from the input tab-delimited text datapack (Fig. A2 of Appendix B) using JavaCC (Java Compiler Compiler) parser and stores the information in various Java data objects.
- (2) **Array of ranges.** Calculate the total number of unique range lines from the parsed data.
- (3) **Tree structure parameters.** Deduce the Parent–Child relationships for each of the range lines to construct the evolutionary-tree data structure. For example, in the horse evolutionary tree (Fig. A2 of Appendix B), the program can parse the information that the *Hyracotherium* parent is branching out to the *Orohippus* child. A typical tree data structure consists of nodes and edges/links from parent nodes to children nodes. In the *TSCreator* evolutionary-tree data structure, a vertical range line refers to a node, and the branch line from a parent range to a child range refers to an edge/link. Each parent range technically can have as many children as required depending on the evolutionary history. A parent range is programmed to keep track of information of its children to determine spacing, coloring, and other attributes in order of geological time. Our evolutionary tree creates a subtree rooted at each child range branching out from the parent range and gradually expands the tree (within the temporal scope of chart, assigned by the Top Age and the Base Age within Settings).
- (4) **Range lines.** Each range line consists of multiple range data points, which together convey a number of flags. For each range line:

- i. Range point information (each item in a single row in the dataset) is extracted. For example, the starting horse range line of North America named “*Orohippus*” has three range points (lower panel in Fig. A2 of Appendix B)
 - ii. Point types are BASE (start of the line), TOP (end of the line), or intermediate branch points shown as “red-dots”. Information for branch labels and branch colors are assigned for each branch point.
 - iii. Drawing style, such as the line width and color, between each pair of range points is assigned. Special flags are set for situations such as whether a point (red-dots) will be blinking if the branch and subtrees are visible or not.
 - iv. Priority flags are set when branch priorities are enabled. Priority values are inherited by higher branches of the tree depending on the value.
- (5) **Branching structure.** This was the most complex suite of algorithms. A left–right branching algorithm partly enables the *TSCreator* evolutionary tree to look visually balanced and semi-symmetric. In the *TSCreator* tree data structure, a range line acts as a parent node, from which left or right child ranges branch. To avoid or minimize possible collisions, the algorithm carefully places child ranges either left or right such that vertical range lines minimally interferes with the horizontal branch lines coming from the parent range. The centering of the parent range and the left/right positioning of each child range follow these simple rules:
- i. Single child — If only a single child, then its range is positioned to the left side of its parent range.
 - ii. Multiple children — If there is an odd number of children, then the earliest (stratigraphically lowest) child is positioned to the left side of the parent, then the next higher child to the right, and this left/right alternation is continued. If there is an even number of children, then the earliest child is positioned to the right side of the parent, and so on. This rule logically matches with the previous rule, because a single child is an odd number.
 - iii. Coeval descent — If two or more children arise from the parent simultaneously, then the left/right positioning of each

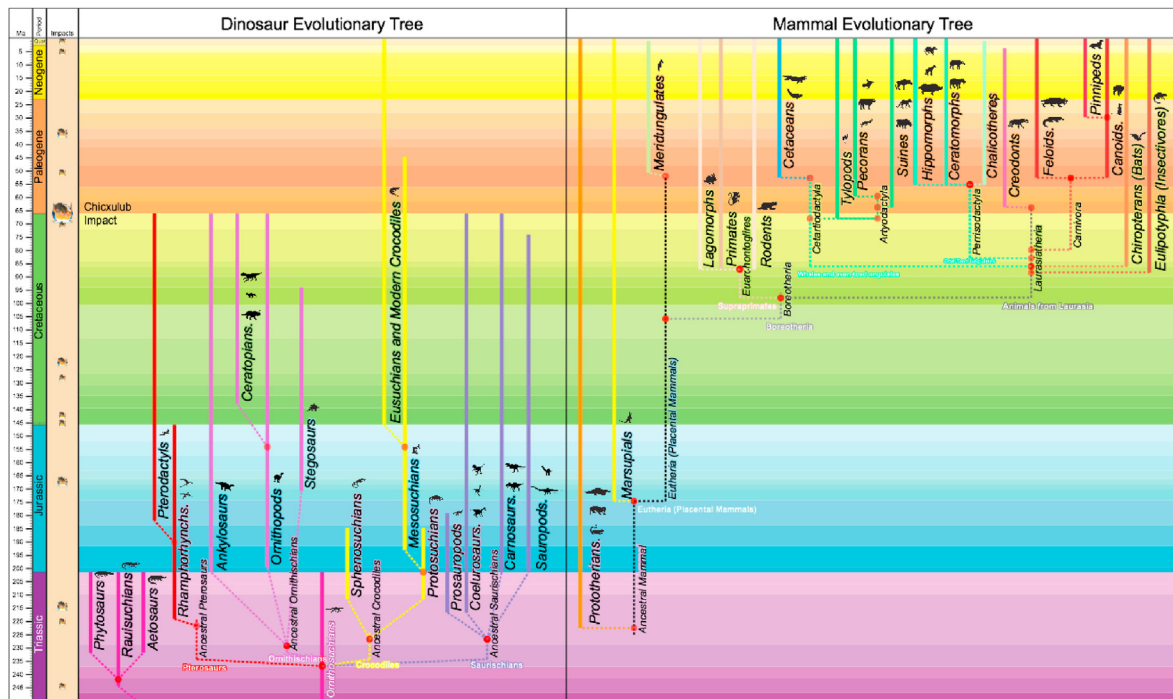


Fig. 7. Using multiple evolutionary and geologic event columns within TSCreator. The datapack on vertebrate evolution was added into TSCreator; then the Dinosaur and the Mammal evolutionary trees were activated, in addition to the internal geologic time scale and major-bolide-impact column. The Dinosaurs (without birds and crocodiles) were extinguished at the Cretaceous/Paleogene boundary (boundary between green and the upper orange colors), followed by the rapid evolutionary divergence of Mammal groups. This major change in the dominant land vertebrates coincides with the Chicxulub impact (the fireball in the tan-shaded left column). The option of background “geologic stage colors” was turned on to highlight the time intervals in the evolutionary tree columns.

child range depends on the order that each child appeared in the text dataset.

The tree branching algorithm then calculates the line height and width of each subtree branching out from a parent range. It keeps track of three main variables: (a) y coordinate of the TOP (LAD: Last Appearance Datum) range point, (b) y coordinate of the BASE (FAD: First Appearance Datum) range point, (c) x coordinate or the horizontal location of the range line. When the parent–child relationship of each range line is established, the x coordinates of a parent and its children are determined programmatically based on the number of children and required widths. All of their widths are calculated simultaneously using recursive function calls (from child to parent). The width of a tree rooted on the parent range is the accumulation of widths of all the subtrees created by its child ranges from both left and right sides. The width determination function takes into account the width of range lines depending on their line styles (e.g., frequent, common, rare, abundant etc.), range labels, pop-up rectangles containing images, and global margin and range padding values provided in the settings window. Width of range-label names again depends on the font size and associated image dimension. All the ranges are finally stored in a single array before drawing where they are, by default, sorted according to the increasing y coordinate value of the BASE range point, i.e. the earlier organisms occur before the later ones. The algorithm first draws the range lines in that sorted order; therefore, the range lines of earlier organisms are drawn on the SVG chart before those that appeared later. The algorithm follows an iterative process (from parent to child) according to the rules a, b, and c to determine the x coordinate of each range line in the evolutionary tree. After drawing the vertical range lines, the algorithm finally connects the child ranges to the parent by drawing the horizontal branch lines using the BASE age, branch point age, and the x coordinate information.

(6) **Color inheritance.** The color is specified in the input dataset in the “BranchColor” field, which the algorithm combines with the

branch line-style information to ultimately determine the drawing style for a horizontal or left/right-inclined branch. Because the algorithm already established the parent–child relationship between the ranges, it can also allow the color inheritance from branch to subsequent child ranges.

- (7) **Space conservation** is a feature in which new evolutionary branches extend into the blank spaces above extinct branches (Fig. 4). This is similar to “new higher branches growing over older lower ones” in a tree and reduces empty space on the chart. Because the ages in the range line interval are already extracted and our algorithm calculates the width of every subtree, then keeping track of empty space above extinct branches is just an extra step. This “Conserve Chart Space in Family Tree Plotting” option is in the main setting menu in TSCreator.
- (8) **Age labels and spacing.** The tree-drawing algorithm simultaneously determines the locations of numerical-age labels while determining the range line positions (e.g., the numerical ages on the ranges of horses in Fig. 3). Age labels are located in the left sides of the ranges and are set to display 2 decimal places. Users can adjust the placement and other paddings using the settings-menu (Fig. 2).
- (9) **Priority.** With this feature certain branches can be prioritized over others when on-screen diagrams become too large (Fig. 5). For instance, high priority could be assigned to organisms that evolve into other major groups, and progressively lesser priority to their side branches. With this feature and the ability by users to turn on/off branch nodes (§ Turning branches on/off, below) in the onscreen display, our TSCreator program has become highly scalable.
- (10) **Turning branches on/off.** Another powerful feature that increases the scalability of TSCreator program to explore large trees is the option for an online user to toggle branch nodes on/off in order to show or hide specific tree branches (Fig. 6). Left-clicking on a “red-dot” node turns off the connected branch. That “red-dot”

blinks on-screen to notify the user that it is turned off and can be activated again by a left-click. The algorithm knows from the pre-established parent–child relationship to change the visibility flag to be false for all the subsequent child ranges; therefore, then recalculates the position/location attribute of the ranges of all other subtrees to redraw the chart. We also provide “Hide Ancestors” option in the settings window; after enabling, you can hide all the parents of a range clicking on a “red-dot” of a range. Selecting “Show All Branches” restores the full tree.

- (11) **Display evolutionary path history and common ancestor.** One can display the evolutionary history of any organism, instead of visualizing the entire evolutionary tree. The tree column options menu provides a textbox to input the name of the organism; and, upon clicking “Show evolutionary history”, a user transforms the full evolutionary tree into a partial tree showing only the evolutionary history of that organism. The interface allows input for two organisms to show the tree rooted at their common ancestor.

3. Additional features and applications of *TSCreator* evolutionary tree display

3.1. Coding groups by a group label and inherited color

In *TSCreator*, we can specify “*BranchColor*”, which is by default inherited by the child ranges and we can also give a “*BranchLabel*”. For example, this label/color option highlights that Ancestral mammals branch into Marsupial (yellow solid range line) and Placental mammals (black dotted range line) during middle of the Jurassic Period (Fig. 7). This feature helps highlight origins and emergence of groups or distinctive taxonomic character sets.

3.2. Display of multiple evolution and data columns

An evolutionary tree can often become quite large and complex. Such a complex chart does provide the opportunity to scan clusters and patterns, but it can quickly become too complicated visually to extract information for specific ranges. However, it is easy to create a datapack with both an overview mega-evolutionary tree and detailed individual evolutionary trees (e.g., our *Tree of Life* datapack; based on Hedges et al., 2009). *TSCreator* allows drawing of multiple evolutionary trees in addition to any number of other columns on Earth history (Fig. 7). This provides an elegant means by which to explore other potential factors and feedbacks relevant to evolution (e.g., bolide impacts in, carbon-isotope trends, etc.).

3.3. Chart exporting and other data-formats

3.3.1. Graphic export formats

Once a chart is constructed with the desired options (fonts, colors, age labels, etc.), it can be exported in established image formats (PDF, SVG, PNG, JPEG). The SVG charts can be enhanced within other graphic-design programs.

3.3.2. Exporting and importing of other tree data formats

The *TSCreator* tree control panel provides an option to export a *TSCreator* tree using standard Newick and Nexus tree formats. This provides a way to transfer our simple tab-delimited *TSCreator* tree data in a format that will load into other software, such as Mesquite (Maddison and Maddison, 2018) or with some R packages to further perform rigorous statistical analysis. See “Instructions to test various tree display features” (Instruction 10) and “Compare *TSCreator* tree output with other software packages” for further details in Supplementary File 1. Fig. 2 in Supplementary File 1 shows an example of how a tree created and visualized in *TSCreator* can also be visualized in Mesquite software package that allows further statistical analysis. However, if the tree data was encrypted by another provider before visualization, then this export

option is disabled, because we respect the interest of the datapack maker to keep their data confidential. Similarly, one can load Newick/Nexus trees exported by other software directly into the *TSCreator* using the “Add Datapack” option under the “File” menu. However, because *TSCreator* is designed to visualize trees associated with geologic time, an arbitrary Newick tree without a deep-time distance measure can be misinterpreted when displayed alongside internal geologic timescales.

3.4. (Data availability) selected examples of current evolutionary tree datapacks for public use in *TSCreator* (at website,⁷ under “datapacks” tab)

- i. **Evolution of Modern life:** This datapack includes evolutionary history from Archea microbes to primates (about 3500 organisms) throughout 4 billion years of Earth history (based on Hedges et al., 2009).
- ii. **Vertebrate Evolution:** About 50 columns of schematic evolutionary relationships from the earliest fish to dinosaurs to primates; accompanied by schematic images and external links.
- iii. **Human Evolution:** Although not a phylogeny, this has main Homo and Australopithecus ranges and tool intervals (ca. past 5 Myr, includes images)
- iv. **Planktonic Foraminifer Evolution through the Paleocene–Eocene:** Dataset and images are from Chenglie Liu et al. (ExxonMobil Exploration) for general release (includes images of 200 taxa).
- v. **Cenozoic Macroperforate Planktonic Foraminifera:** Phylogenetic tree data (Aze et al., 2011) updated to GTS2016 with images (Fordham et al., 2018). Detailed morphospecies evolution is shown alongside macroevolutionary lineage trees.

4. Conclusion

Most current evolutionary-tree graphical packages lacked the capability to display trees against other time-scaled geologic or geochemical data, or they do not enable click-access to the underlying data through pop-up windows. The embedding of an evolutionary-tree display with an array of special features within the *TimeScale Creator* platform enables paleontologists and other geoscientists, biologists and science students to visually compare the evolutionary events to other biologic, lithologic, magnetic, sea level, chemical and tectonic records from different regions to investigate potential relationships. Our addition of evolutionary-tree graphics to the *TimeScale Creator* visualization software for Earth history provides a powerful and flexible method to display and study development of life through deep time.

Link to code

- (1) The entire *TimeScale Creator* (JAVA) – free at <https://timescalecreator.org>. Both.jar (for Mac) and.exe (for Windows) are free to download, and there are extensive user manuals for all aspects, not just the evolutionary tree. In addition to the internal datapack (over 20,000 events in Earth History), there are about 20 additional datapacks contributed by geologic surveys (Canada, Britain, Australia, etc.) or compiled from publications by our students (including *Tree of Life*, from Hedges and Kumar).
- (2) The source code for all routines of the evolutionary-tree package development/enhancement through the public github repository: https://github.com/brishtiteveja/TSCreator_Evolutionary_Tree_Visualization

Authorship statements

A. K. Zehady, a Purdue University graduate student working on data

⁷ <https://timescalecreator.org/>.

mining of Earth history databases in geoscience and computational science (now at Ciscos), programmed the evolutionary tree software and authored the drafts of the manuscript. J. G. Ogg, now on retired status at Purdue University and a distinguished visiting professor at Chengdu University of Technology, supervises the *TSCreator* program and co-wrote the final manuscript. B. G. Fordham, a retired adjunct at Australian National University, was instrumental guiding the development of evolution visualization within the *TSCreator* system and compiled the planktonic foraminifer evolution database for this project. G. Palem, a graduate student from Purdue University in computer science and now a software engineer at Google, developed the initial version of evolution display. J. Bobick, a master's student in computer and information technology at Purdue University, technical assistant for *TimeScale Creator* software project. G. Ogg, a paleontologist (PhD from Tübingen), coordinated testing and user-interfaces for the display of evolutionary trees, prepared the final versions for the *TSCreator* website and enhanced the graphics.

Declaration of competing interest

The authors declare that they have no known competing financial

Appendix A. *TSCreator* datapack format for ranges

In the range-column dataset, the initial tab-delimited header (green-shaded rows in Fig. A1) includes the title for the display (column A) and the flag “range” (column B) to indicate the column type. Optional fields are a background color in R/G/B (column D), whether the title should be displayed (column E), whether the column should be automatically on or off (column F), and a comment field (column G) that is activated as a pop-up window when the column title is mouse-clicked.

This header is followed by the range data for each successive organism (tan-shaded rows in Figure A1). The entry consists of the name (column B), the (earliest) age to begin a line style (column C), the type of line style (column D; see left side of Fig. 2 for the types), an optional comment field (column E), and an optional line color (Column F). The line types (column D) are merely flags to set a line thickness or style, except that “TOP” terminates a line, “missing” is a flag to insert a gap, and “sample” superimposes a dot on any portion of a range-line or outside of that range. During the loading of the dataset, the program looks for all data with the same name, and then draws the indicated line style upward (higher/older) until either an age for different line style is indicated, the word “TOP” is indicated in the line-style field, or the top of the selected age span is encountered.

A	B	C	D	E	F	G	H
<Title>	range	<blank>	<color>	notitle	on/off	<popup>	Header Row
<blank>	<label>	<age>	<abundance>	<popup>			Data Row

Example

A	B	C	D	E	F	G
Dino-cysts	range		175/201/108	notitle	off	fictitious dino abundances
	Dracodinium varielongitutum	50.56	TOP	you could also add an URL link into this popup		
	Dracodinium varielongitutum	51.3	rare	poorly preserved		
	Dracodinium varielongitutum	52.6	abundant			
	Dracodinium varielongitutum	53.44	frequent			

⁸ <https://timescalefoundation.org>.

Fig. A1. A portion of the dataset for the dinoflagellate range chart of Fig. 2 in Excel sheet format; see text for description of header (green shaded) and data (tan shaded) fields. The name “*Dracodinium varielongitudum*” (column B) will display vertically along its range (leftmost range line in Fig. 2). That displayed name can include an embedded image using an HTML image tag that provides the alignment, scaling and image file name. In this example, the thumbnail picture was pre-rotated 90° so it will look upright when the range label is plotted vertically (see Fig. 2). Values in Column C are ages in Ma for the beginning of each line style within the range line (Column D) with optional pop-up information (Column E). To insert a URL link within the pop-up text, one uses HTML anchor tag format – e.g., including “[Click Adnatosphaeridium_multispinosum for details]” will result in an active web link from “Adnatosphaeridium_multispinosum”. If Excel is used to enter the dataset, then the sheet needs to be exported as a tab-delimited text file and merged with referenced.PNG or.JPG images into a zipped file, for loading into the TSCreator program.

Appendix B. TSCreator datapack format for evolutionary tree

A “branch” flag consisting of four items is all that is required to enhance the format coding of range data in order to display connections among ranges for an evolutionary tree display in TSCreator: the name of the Parent range (Column B in Fig. A2), the age of branch from its range (C), the flag “branch” (D) and the name of the Child range (E). Other display options have been progressively added to that basic “branch” statement:

Parent-name (B) ⇒ Age (Ma) (C) ⇒ “branch” (D) ⇒ Child-name (column E) ⇒ “on/off” (column F) ⇒ BranchLabel (G) ⇒ Dashed/Dotted (H) ⇒ Pop-up (I) ⇒ BranchColor (J) ⇒ Priority (K)(the “⇒” between the items below indicates a [TAB], and the letters inside brackets indicate the column in the Excel sheet of Figure A2).

Datapack Format for Evolutionary Tree Columns

A	B	C	D	E	F	G	H	I	J	K	L	
<Title>	range	<blank>	<color>	notitle	on/off	<popup>						Header Row
<blank>	<label>	<age>	<abundance>	<popup>								Data Row for range
<blank>	<parent name>	<age>	branch	<child name>	on/off	<branch label>	dashed/dotted	<popup>	<branch color>	<priority>		Data Row for branch

Example

A	B	C	D	E	F	G	H	I	J	K
Horses	range		245/232/217		on	the data is not real				
	Hyracotherium	45.39		popup for Hyracotherium						
	Hyracotherium	51	branch	Orohippus		North America	dashed	70% up in Ypresian	144/139/250	10
	Hyracotherium	55.96	frequent	Base = Ypresian						
	Orohippus	45.39		only few fossils						
	Orohippus	48	branch	Epihippus			dashed	Base Lutetian		10
	Orohippus	50.49	frequent	Base = 2/3 up in Ypresian						
	Epihippus	45	TOP	Base = Early Lutetian						
	Epihippus	47.5	flood							
	Hyracotherium	51	branch	Eurohippus		Europe	dashed	70% up in Ypresian	250/0/0	10
	Eurohippus	44								
	Eurohippus	50.6		Base = 70% up in Ypresian						

Fig. A2. Sample dataset for the schematic evolution of horses (used in Fig. 4) in Excel sheet format. Green-shaded rows are header information, tan-shaded rows are range data, and gray-shaded rows are the evolutionary-branch statements. The upper panel is a brief explanation of the fields in the lower panel. See text for details.

The field “*Child-name*” (E) must be identical to the range-name of the Child (but TSCreator ignores an image entry appended to the name). The branch age must be within the age range of the “Parent”; and, obviously, cannot be older than the base of the “Child” range.

The “on/off” (F) option denotes whether the child and its subtree (clade) is to be automatically displayed. If this field is blank, then the default is “on”.

BranchLabel (G) is a horizontal label in larger font for the new subtree and is printed under the branch line coming from the parent range (e.g., “North America” and “Europe” in Fig. 4).

Dashed/dotted (H) provide alternative options to the default solid branch line style.

Pop-up (I) is the content that will be displayed in a pop-up when one left-clicks the selected red-shaded rectangles on range labels or right clicks the red-dot node (Fig. 1). It is our practice to always try to include the calibration of the geologic age and the source of the information. Those pop-ups can also include images and URL links to external websites (Fig. A1).

BranchColor (J) is an RGB value that will be applied to the “BranchLabel” font and to the range lines of the subtree. For example, in Fig. 4 the branch label “North America” is given bluish-gray color and the child range lines of *Orohippus* and *Epihippus* inherit that color. That color will be overwritten if another color is specified for a subsequent branch (and its subtree).

Priority (K) provides the ability to prioritize certain branches over others when on-screen diagrams become unwieldy large (see § Programming the Evolutionary tree display, Priority).

Appendix C. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.acags.2020.100037>.

Computer code availability

TimeScale Creator evolutionary tree software module was developed at Purdue University by Abdullah Khan Zehady, Jason Bobick (A.K. Zehady at 1-765-772-8170, now at Cisco; azehady@purdue.edu) and Gangi Palem (now at Google Inc., Mountain View, CA; palemgangireddy@gmail.com); or contact our mentor, professor James Ogg (jogg@purdue.edu; 1-765-743-0400). The software can be launched in Windows, Macintosh or Linux operating systems. Instructions to run and generate charts using datapacks are in Supplementary file 1. The code is in Java programming language and Java 1.6 or higher version is recommended to execute the program. The code for the evolutionary-tree column portion (only java files; ~500 KB) is freely available under the MIT open source license from a github repository.⁹ The repository also has a ~33 Mb java.jar and a.exe file to directly run the program and generate charts. The software package and all datasets (encoded) are available free of charge from the downloads section of our project homepage¹⁰. The Public *TSCreator* version has no restriction to importing range-type or evolutionary tree columns in tab-delimited *TSCreator* format, in addition to a moderate number of additional self-made columns. There is no limit to the size of range/evolution columns, including pop-up information. However, if additional non-range/evolution columns consist of more than 3000 characters, then a Pro license should be requested (which is usually free for students). In that case, if you are an academic or a researcher and are willing to make your larger dataset publicly available through our website, a free Pro version is also possible.

References

- Alix, B., Boubacar, D.A., Vladimir, M., 2012. T-REX: a web server for inferring, validating and visualizing phylogenetic trees and networks. *Nucleic Acids Res.* 40 (W1), W573–W579. <https://doi.org/10.1093/nar/gks485>.
- Applied Maths, 2020. BioNumerics SEVEN: a Unique Software Platform. Software for Analysis of Electrophoresis Patterns, Phenotype Arrays, Sequences and Much More. <http://www.applied-maths.com/bionumerics>.
- Aze, T., Ezard, T.H., Purvis, A., Coxall, H.K., Stewart, D.R., Wade, B.S., Pearson, P.N., 2011. A phylogeny of Cenozoic macroperforate planktonic foraminifera from fossil data. *Biol. Rev.* 86 (4), 900–927. <https://doi.org/10.1111/j.1469-185X.2011.00178.x>.
- Bobick, R., Ogg, J.G., Salgado Jauregui, E., Santos Filho, M.A.B., Ault, A., students of Purdue University, 2016. Earth History Visualization System for Gondwana Basins. 35th International Geological Congress (27 Aug – 4 Sept, 2016, Cape Town, South Africa) Abstracts. American Geoscience Institute. <https://www.americangeosciences.org/igc/16592>. <https://www.americangeosciences.org/sites/default/files/igc/5196.pdf>.
- Cazaux, B., Castel, G., Rivals, E., 2019. AQUAPONY: visualization and interpretation of phylogeographic information on phylogenetic trees. *Bioinformatics* 35 (17), 3163–3165. <https://doi.org/10.1093/bioinformatics/btz011>.
- Chen, W.C., 2019. Phyclus: Phylogenetic Clustering (Phyloclustering). Iowa State University, Ames, IA, USA. <https://cran.r-project.org/web/packages/phyclus/>.
- Commission for the Geological Map of the World, 2012. Stratigraphic chart: paris [cited 2018 [color codes provided for various products, see also appendix in Gradstein et al., 2012]]. <https://ccgm.org/en/27-stratigraphic-chart>.
- Drummond, A.J., Rambaut, A., 2007. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol. Biol.* 7 <https://doi.org/10.1186/1471-2148-7-214> article #214.
- Felsenstein, J., 2019. PHYLIP (phylogeny inference package), version 3.698 c. Joseph Felsenstein. last update was 2014. <http://evolution.genetics.washington.edu/phyli.html>. (Accessed 15 April 2020).
- Felsenstein, J., 2014. Phylogeny Programs last update was 2014. <http://evolution.genetics.washington.edu/phyli/software.html>. (Accessed 15 April 2020).
- Fordham, B.G., Aze, T., Haller, C., Zehady, A.K., Pearson, P.N., Ogg, J.G., Wade, B.S., 2018. Future-proofing the Cenozoic macroperforate planktonic foraminifera phylogeny of Aze & others (2011). *PLoS One* 13 (10), e0204625. <https://doi.org/10.1371/journal.pone.0204625>.
- Francisco, A.P., Vaz, C., Monteiro, P.T., Melo-Cristino, J., Ramirez, M., Carriço, J.A., 2012. PHYLOViZ: phylogenetic inference and data visualization for sequence based typing methods. *BMC Bioinf.* 13 <https://doi.org/10.1186/1471-2105-13-87> article #87.
- Gradstein, F.M., Ogg, J.G., 2006. Chronostratigraphic database and visualization; Cenozoic-Mesozoic-Paleozoic integrated stratigraphy and user-generated time scale graphics and charts. *GeoArabia* 11 (3), 181–184. <https://doi.org/10.6084/m9.figshare.11872155>.
- Gradstein, F.M., Ogg, J.G., Schmitz, M.D., Ogg, G.M., 2020. The Geologic Time Scale 2020. Elsevier, 1400 pp. <https://www.elsevier.com/books/geologic-time-scale-2020/gradstein/978-0-444-63798-7>.
- Guindon, S., Dufayard, J.F., Lefort, V., Anisimova, M., Hordijk, W., Gascuel, O., 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst. Biol.* 59 (3), 307–321. Software. <http://www.atgc-montpellier.fr/phyml/>.
- Hedges, S.B., Kumar, S. (Eds.), 2009. The TimeTree of Life. Oxford University Press, 551 pp. <http://timetree.org/book>.
- Helaers, R., Milinkovitch, M.C., 2010. MetaPIGA v2. 0: maximum likelihood large phylogeny estimation using the metapopulation genetic algorithm and other stochastic heuristics. *BMC Bioinf.* 11 <https://doi.org/10.1186/1471-2105-11-379> article #379.
- Huelsenbeck, J.P., Ronquist, F., 2001. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* 17 (8), 754–755. <https://doi.org/10.1093/bioinformatics/17.8.754>.
- Huson, D.H., Richter, D.C., Rausch, C., DeZulian, T., Franz, M., Rupp, R., 2007. Dendroscope: an interactive viewer for large phylogenetic trees. *BMC Bioinf.* 8 (1) <https://doi.org/10.1186/1471-2105-8-460> article #460.
- Kreft, L., Botzki, A., Coppens, F., Vandepoole, K., Van Bel, M., 2017. PhyD3: a phylogenetic tree viewer with extended phyloXML support for functional genomics data visualization. *Bioinformatics* 33 (18), 2946–2947. <https://doi.org/10.1093/bioinformatics/btx324>.
- Letunic, I., Bork, P., 2006. Interactive Tree of Life (iTOL): an online tool for phylogenetic tree display and annotation. *Bioinformatics* 23 (1), 127–128. <https://doi.org/10.1093/bioinformatics/btl529>.
- Liu, L., Yu, L., 2010. Phybase: an R package for species tree analysis. *Bioinformatics* 26 (7), 962–963. <https://doi.org/10.1093/bioinformatics/btq062>.
- Maddison, W.P., Maddison, D.R., 2018. Mesquite: a modular system for evolutionary analysis. Current release 3, 61. <http://www.mesquiteproject.org>.
- Meade, A., Pagel, M., 2019. BayesTraits. Reading evolutionary biology group. Current release 3, 0.2 (Nov., 2019). <http://www.evolution.rdg.ac.uk/BayesTraitsV3.0.2/BayesTraitsV3.0.2.html>.
- Milne, I., Lindner, D., Bayer, M., Husmeier, D., McGuire, G., Marshall, D.F., Wright, F., 2009. TOPALi v2: a rich graphical interface for evolutionary analyses of multiple

⁹ https://github.com/brishriteveja/TSCreator_Evolutionary_Tree_Visualization.

¹⁰ <https://timescalecreator.com>.

- alignments on HPC clusters and multi-core desktops. *Bioinformatics* 25 (1), 126–127. <https://doi.org/10.1093/bioinformatics/btn575>.
- Ogg, J.G., Ogg, G.M., Gradstein, F.M., 2016. Concise Geologic Time Scale 2016. Elsevier Publ. <https://doi.org/10.1016/c2009-0-64442-1>, 234 pages.
- Ogg, J., Lugowski, A., Gradstein, F., 2010. Earth history databases and visualization – the TimeScale creator system. In: EGU General Assembly Conference Abstracts (2-7 May 2010, Vienna). Available at: <https://ui.adsabs.harvard.edu/abs/2010EGUGA..12.7039O/abstract>.
- Ogg, J., Przybylski, P., 2009. Jurassic chronostratigraphic database and the *TimeScale Creator* visualization system. *Volumina Jurassica* 7, 175–179. <https://vjs.pgi.gov.pl/article/view/26548/18252>.
- Okonechnikov, K., Golosova, O., Fursov, M., the UGENE team, 2012. Unipro UGENE: a unified bioinformatics toolkit. *Bioinformatics* 28 (8), 1166–1167.
- Revell, L.J., 2012. phytools: an R package for phylogenetic comparative biology (and other things). *Methods in Ecology and Evolution* 3 (2), 217–223. <https://doi.org/10.1111/j.2041-210X.2011.00169.x>.
- Revell, L.J., Chamberlain, S.A., 2014. Rphylip: an R interface for PHYLIP. *Methods in Ecology and Evolution* 5 (9), 976–981. <https://doi.org/10.1111/2041-210X.12233>.
- Robinson, O., Dylus, D., Dessimoz, C., 2016. Phylo.io: interactive viewing and comparison of large phylogenetic trees on the web. *Mol. Biol. Evol.* 33 (8), 2163–2166. <https://doi.org/10.1093/molbev/msw080>.
- Schliep, K.P., 2010. phangorn: phylogenetic analysis in R. *Bioinformatics* 27 (4), 592–593. <https://doi.org/10.1093/bioinformatics/btq706>.
- Shank, S.D., Weaver, S., Koskovsky Pond, S.L., 2018. phylotree.js - a JavaScript library for application development and interactive data visualization in phylogenetics. *BMC Bioinf.* 19 <https://doi.org/10.1186/s12859-018-2283-2> article #276.
- Smith, T., Ogg, J.G., Kelman, A., Abbott, S., Bernecker, T., 2015. Toward updated stratigraphic frameworks for Australia's offshore hydrocarbon provinces. *The Australian Petroleum Production & Exploration (APPEA) Journal* 55 (1), 105–111. <https://doi.org/10.1071/AJ14008>.
- Smits, S.A., Ouverney, C.C., 2010. jsPhyloSVG: a javascript library for visualizing interactive and vector-based phylogenetic trees on the web. *PLoS One* 5 (8), e12267. <https://doi.org/10.1371/journal.pone.0012267>.
- Stöver, B.C., Müller, K.F., 2010. TreeGraph 2: combining and visualizing evidence from different phylogenetic analyses. *BMC Bioinf.* 11 (1) <https://doi.org/10.1186/1471-2105-11-7> article #7.
- Vaughan, T.G., 2017. IcyTree: rapid browser-based visualization for phylogenetic trees and networks. *Bioinformatics* 33 (15), 2392–2394. <https://doi.org/10.1093/bioinformatics/btx155>.
- Yu, G.C., Smith, D.K., Zhu, H.C., 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>.
- Zhang, H.K., Gao, S.H., Lercher, M.J., Hu, S.N., Chen, W.H., 2012. EvolView, an online tool for visualizing, annotating and managing phylogenetic trees. *Nucleic Acids Res.* 40 (W1), W569–W572. <https://doi.org/10.1093/nar/gks576>.