Treevolution

User's Guide

Version 1.2

July 20, 2009

VNiVERSiDAD Ð SALAMANCA

VIS

u s a l

Rodrigo Santamaría and Roberto Therón
Departamento de Informática y Automática

Contents

List of Figures

Chapter 1

Installation

1.1 License

Treevolution is distributed under the following license:

The MIT License Copyright (c) 2009 Universidad de Salamanca
Permission is hereby granted, free of charge, to any person obtaining a copy of this
software and associated documentation files (the "Software"), to deal in the
Software without restriction, including without limitation the rights to use, copy,
modify, merge, publish, distribute, sublicense, and/or sell copies of the Software,
and to permit persons to whom the Software is furnished to do so, subject to the
following conditions:
The above copyright notice and this permission notice shall be included in all
copies or substantial portions of the Software.
THE SOFTWARE IS PROVIDED "AS IS", WITHOUT WARRANTY OF ANY
KIND, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO THE
WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR
PURPOSE AND NONINFRINGEMENT. IN NO EVENT SHALL THE
AUTHORS OR COPYRIGHT HOLDERS BE LIABLE FOR ANY CLAIM,
DAMAGES OR OTHER LIABILITY, WHETHER IN AN ACTION OF
CONTRACT, TORT OR OTHERWISE, ARISING FROM, OUT OF OR IN
CONNECTION WITH THE SOFTWARE OR THE USE OR OTHER
DEALINGS IN THE SOFTWARE.

1.2 Requirements

Treevolution is a software developed in Java 1.6 with the help of the Processing library [1][1]. It is expected to work under any operating system that support the required JRE version[2]. Treevolution requires:

- JRE 1.6 or higher[3].

- 20MB of disk space.

- 1GB of RAM (more RAM is recommended to visualize trees of more than one thousand nodes).

1.3 Installation and Run

To obtain Treevolution, go to *http://vis.usal.es/treevolution* and take the latest version in the *downloads* section. Please note that alpha versions can still have some bugs. After downloading the package, please follow these steps:

1. Unzip the downloaded file (*treevolutionX-X.rar*) wherever you want to install the software

2. Open the unzipped folder

3. Depending on the OS:

    - *Windows:* Double click on file *treevolution.bat* or, if a command line is used, go to the corresponding folder and write *treevolution.bat*
    - *Linux:*
        - Check that *treevolution.bat* has execution privileges. If not, give them using *chmod +x treevolution.bat.*
        - Run the application: *./treevolution.bat.*

    If everything is OK, Treevolution should start right away. If you have any question or problem, please contact us at rodri@usal.es.

---

[1]The stand-alone Processing can be downloaded at *http://processing.org*, but Treevolution comes with an integrated Processing package, so it is not required.

[2]Note that, as of now, this excludes MacOS X 32-bits version

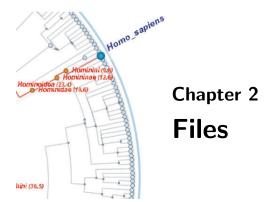[3]You can download it at *http://java.com*

1.3.1 Screen size

Treevolution is designed to be displayed in the full screen, without window frames. If the ratio screen width/screen height is above 1.8, then only half the width is occupied (as, for example, in 2-monitor configurations). If you want to set up your own size limits, you can change the arguments of Treevolution by editing the treevolution.bat file. This is the usage:

```
java -Xmx1024m -jar bin/te.jar [false|true] [width] [height]
```

[false—true] specified if Treevolution is decorated with the operating system window frame (true) or not (false). Default is false.

[width] and [height] determine the size of the Treevolution window in pixels. By default, it fits the full screen size.

Chapter 2

Files

2.1 File Formats

Treevolution can load trees in Newick (.nwk) or PhyloXML (.xml) formats. New Hampshire eXtended format (.nhx) is to be migrated to PhyloXML format, so only PhyloXML format is supported by Treevolution[4]. Note that Treevolution is originally designed to deal with species trees. Other kinds of trees, such as statistically inferred trees with confidence or bootstrap branching values, etc. can be properly read by Treevolution (provided that their actual format is correct), but it will only convey names and branch lengths. An appropriate representation, in line with the rest of the Treevolution design choices, for the rest of characteristics is scheduled for the next version.

The *Newick* format is the simplest one, with just node names and branch lengths. The extended properties of the .nhx format are not supported.

The *PhyloXML* format is fully supported, and has been tested with the online examples at *http://phyloxml.org*. The *name*, *taxonomy*, *node_id*, *branch_length*, *events* and *sequence* tags are conveyed in the representation in the current version:

- Node labels are *node_ids* (extracted from the input file if they exist, otherwise they are automatically generated), unless there are *name* or *taxonomy* labels associated to the clade. If both are present, the label is "name/taxonomy".

- Clades (nodes) without *branch_length* tags have a branch length of 1[5]. Attached to node labels, there is a number in parenthesis that indicates the total branch length from the current node to the most distant descendant. Note that this nomenclature fits with species trees that have time information as branch lengths (chronograms), but may not be the best representation for other kinds of trees.
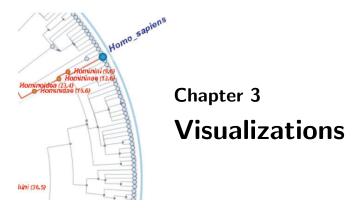
---

[4]You can find a NHX to PhyloXML format converter at:
http://phylosoft.org/forester/applications/phyloxml_converter

[5]Note that Treevolution is designed to divide the radial dendrogram into rings at distance 1

- The *events* tag determines the color of the point of branching. If the event is
  an speciation, it is green, if it is a duplication, it is red. Otherwise, it is grey.

- The *sequence* must determine a *domain_architecture* (see the *data/apaf.xml*
  file for example). In this case, the domains are represented in different colors
  depending on their name, in a figure besides the node label. Only leaf nodes
  can have a sequence associated.

Treevolution's distribution contains some example trees in the *data* folder.
Newick data files come from [2] supplementary material. The PhyloXML data files
are the examples available online at PhyloXML web site.

Chapter 3

Visualizations

Treevolution supports three visualizations:

- *Radial dendrogram*: visualizes the tree on a circular display. Sectors identify families and rings depth. It allows to represent more leaf nodes than a linear dendrogram, and implements several interaction techniques to explore the tree.

- *Linear dendrogram*: visualizes the tree or sections of the tree on a linear display. It is designed to represent small sections of the tree.

- *Histograms*: bar charts that display simple statistics about the tree, such as number of descendants per node, the number of nodes within each ring, etc.

The user can switch between them with the top right tabs (see fig. 3.1).

Figure 3.1. *Interface to switch among visualizations (from top to bottom: radial dendrogram, linear dendrogram and histograms). The linear dendrogram is disabled until a node is locked by right clicking on it (the top lock icon indicates wether there is a locked node or not).*

3.1   Radial dendrogram

This is the main visualization and represents the whole tree. A tree usually has much more leaf nodes than internal nodes, and a circular representation exploits this situation, enabling the representation of larger trees (see fig. 3.2). The radial dendrogram can be rotated by using the button on the top of the circular representation. Treevolution's radial dendrogram is divided into rings and sectors.
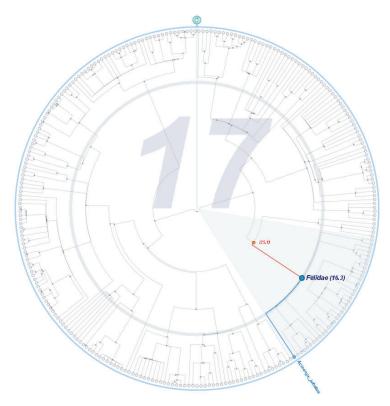

Figure 3.2. *Radial dendrogram with the* Felidae *node selected. The ring it pertains to (17) and the sector wrapping its family are highlighted.*

3.1.1   Rings

Rings convey one unit steps in branch lengths. For example, if branch lengths covey evolution times in terms of millions of years (Myrs), each ring represents 1 Myr. For more abstract tree branches, such as those of automatically generated trees, the meaning intervals is more artificial, but still are a good way to delimit zones in the tree.

Each ring is labeled with a number representing the depth from the periphery, rounded up. This number appears in the background of the dendrogram. By pressing *F5* function key a circumference is drawn for each ring (otherwise, ring

boundaries are hidden if they are not hovered over). Rings may be distorted and filtered in order to focus on determinate sections of the dendrogram.

Manual distortion

The simplest way to focus on a given ring is to manually increase its size by dragging the boundary of the ring (see fig. 3.3). The width of rings can be reset to their initial sizes with the corresponding button (see chapter 4).
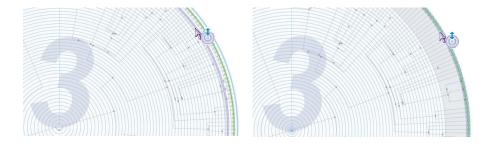

Figure 3.3. *By dragging the mouse from a ring boundary, the user can modify ring widths.*

Fisheye distortion

Fisheye distortion [3] increases the size of the current ring and their neighbors, gradually decreasing the distortion factor (see fig. 3.4). The recommended use of this distortion is to activate it, hover over the dendrogram and whenever the distortion is satisfactory, deactivate it using the *F2* function key. The amount of fisheye distortion can be selected using the proper button (see chapter 4 for a summary of buttons).
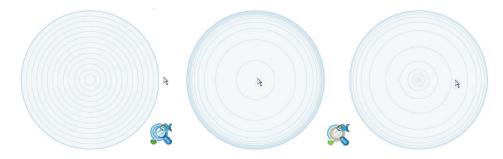

Figure 3.4. *If the fisheye distortion is set, rings can be distorted by hovering over them. Fisheye distortion can be activated or deactivated with the corresponding interface button (the icon is shown in the figure) or with the* F2 *function key.*

Ring Filtering

Finally, depending on the focus of the current exploratory analysis, uninteresting rings may be hidden by means of filtering. If selected, an interval bar can be used to set the ring interval to show (see fig. 3.5).
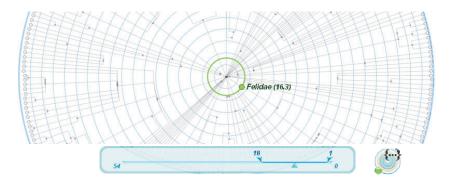

Figure 3.5. *Ring filtering is performed by choosing the ring interval to visualize. In the figure, the period between 18 Myrs. and 1 Myr. ago has been selected, the time frame where different* Felidae *species branched.*

3.1.2   Sectors

Sectors are angular zones surrounding all the descendants from a given element. Initially, there is only one 360 degrees sector from the root node. When a node is hovered over, its corresponding sector is highlighted. In addition, the user may "cut" a given sector, splitting the initial sector and giving this new sector half the original space. This can be done any number of times, each one giving the new sector half the space of the sector it was in. The size of the sector can also be modified by dragging its boundaries, or may be removed from the visualization (see fig. 3.6).

Figure 3.6. *The sector starting at* Felidae *has been split by selecting the* Felidae *node and clicking on the corresponding button (left), then the sector size is modified by dragging the sector boundary (center) and finally the rest of the tree is removed by selecting the root node and clicking on the corresponding button (right).*

3.1.3   Node exploration

When a node is hovered over, its corresponding label is shown. The label contains the node name and taxonomy (if available), plus a number in parenthesis with its distance in branch length to the most external leaf node. If the node does not have neither name nor taxonomy, the node id[6] and the branch length are displayed. If it is an internal node, the label is displayed horizontally at the side of the node, if it is a leaf node it is displayed radially at the periphery. A button in the interface allows to display all the labels corresponding to leaf nodes at once. If the density of nodes at a point clutters the visualization (making their labels unreadable), some of the labels are automatically hidden. If a node is clicked, it remains selected, displaying its label and the circumference at its level (see fig. 3.7).
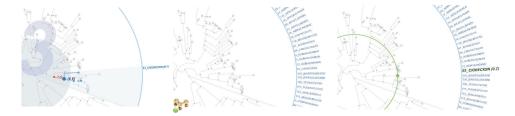

Figure 3.7. *When a node is hovered over, its label is drawn, in this case just informing about its total branch length, 1.1 (left). All the leaf node labels can be drawn (center). If a node is selected, its label and a concentric circumference passing trough it remain drawn in green (right).*

Nodes can be searched by name by using the column at the left (see fig. 3.8).

---

[6]If the *node_id* tag is present in the file, it is used as identifier, if not, or if it is a .nwk file, Treevolution assigns a unique number to each node

The search returns the names of the nodes that contain (ignoring cases) the characters written in the search box at the bottom. The user can select a name in the search column, that in turn selects the corresponding node in the dendrogram.
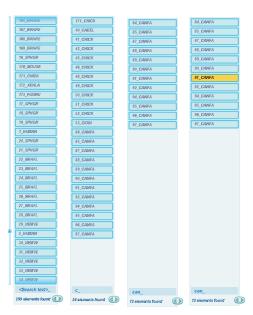

Figure 3.8. *Progressive search for the text* can.

When a node is hovered over, its direct ancestor and descendants are also highlighted, at a smaller size (the ancestor in red, the descendants in blue, see fig. 3.7, left). The interface allows to highlight the indirect ancestors and descendants too (see fig. 3.9).

Finally, right-clicking on a node activates the linear dendrogram visualization, that will display the corresponding highlighted branches (corresponding to the direct or all the ancestors/descendants, depending on the option selected).

3.1.4 Coloring families

By right-clicking on a ring, Treevolution detects every node without ancestors after that point. Then, it colors all the branches coming from each of these nodes (see fig. 3.10). This can be defined as a kind of "visual clustering", helping to give an overall idea of groups and group evolution along rings.
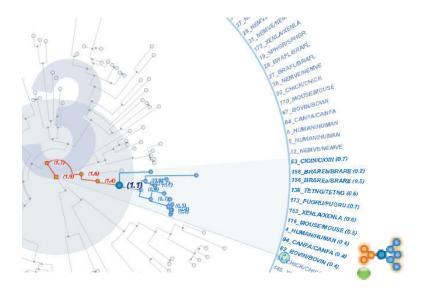
Figure 3.9. *Every direct or indirect ancestor or descendant of the node hovered over in fig. 3.7 (left) is now highlighted.*
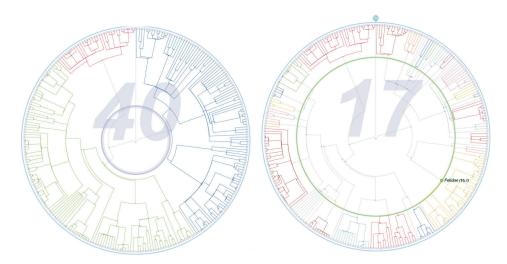

Figure 3.10. *The coloring of carnivore mammals at 40 Myrs. (left) reveals that only three subspecies were present at that time. When* Felidae *branched out at 17 Myrs. (right), there were several species, but note that one of the three families at 40 Myr. (in red, at the top left) still hasn't branched: it is the ancestor of* Canidae.

3.2 Linear dendrogram

This visualization is the linear counterpart of the radial dendrogram (see fig. 3.11). Here, sectors become horizontal areas and rings become vertical columns. Besided

the layout, the interface is analogous, including column distortion (fisheye and manual), column filtering, selection, hovering, labels and text search. There are also some particular interactions such as the rotation of labels 45 degrees in order to avoid cluttering with horizontal branches, and the possibility of moving nodes by right clicking and dragging the mouse.
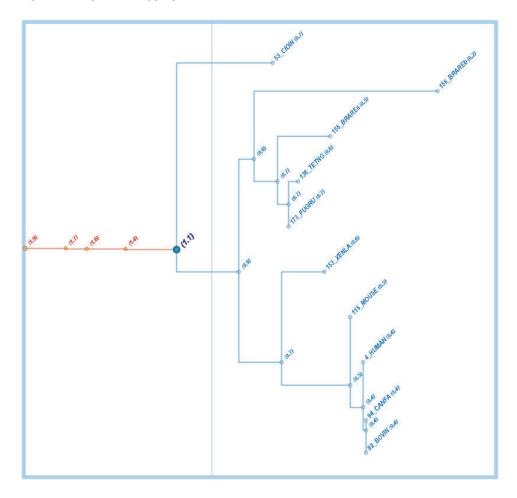

Figure 3.11. *Representation of the branches highlighted in fig. 3.9. Only the last to rings (columns here) are considered (the rest have been filtered out), and the labels have been rotated 45 degrees.*

This visualization is designed to display small subtrees selected in the radial dendrogram. Therefore, to activate this visualization the user must select a node with by right clicking on it. This sets to green the top lock icon. The branches highlighted at the dendrogram are the ones to be represented in the linear dendrogram, so if you want to display the full subtree you must select the option to highlight indirect relatives with the corresponding interface button or the *F3* key.

3.3   Histograms

In Treevolution, histograms represent tree statistics[7], specifically:

- *Number of direct descendants per node.*

- *Number of direct ancestors per node*: although not useful for trees (the number of direct ancestors is one) this option is here in order for future versions to support other kind of graphs with nodes that can have more than one direct ancestor.

- *Number of nodes per ring*: this is specially useful if branch lengths convey evolutionary times, because it visualizes active time frames.

- *Number of nodes per family*: if there is more than one sector (split by the user), this histogram represents the number of nodes within each sector.

The radial dendrogram is displayed in the background of the histogram, and when a bar in the histogram is hovered over or selected, the corresponding element in the dendrogram (node, sector or ring) is highlighted (see fig. 3.12), so the context of the exploratory analysis is not lost.

Figure 3.12. *Representation of the number of nodes per ring on the evolutionary tree for carnivore mammals. The bar for 28 Myrs. is selected, and the corresponding ring is highlighted in the background dendrogram.*

---

[7]Note that all the histograms are in logarithmic scale

Chapter 4

Data Interaction

There are five basic kinds of interaction with the elements in the visualization:

- *Hover*: using use the mouse to hover over an element (either a node, bar or button) highlights it and displays the text information about the element (node name and taxonomy, bar name and value, or button description). To hover over areas (either rings, sectors or columns) shades them with a different color.

- *Left click*: selection of elements by left-clicking the mouse on them usually has the same effect than hovering over, but the effect persists when the mouse changes to other element. To click again on a selected element removes the selection. In order to keep rings, sectors or columns selected you must set first the corresponding option with the proper interface button.

- *Right click*: right-clicking on nodes of the radial dendrogram activates the linear dendrogram for the corresponding family. Right-clicking on a ring of the radial dendrogram colors branches depending on their common ancestor at the depth level of the clicked ring.

- *Drag*: Ring, sector and column boundaries can be dragged to modify their sizes. Nodes in the linear dendrogram can be dragged up and down, by using the right button.

- *Text search*: The left column at the left of dendrograms allows text searches by name. To select labels in this column selects the corresponding node in the dendrogram.

Apart from direct interactions with elements of the visualization, there are several additional options to modify the views. Most of them have been detailed in the former chapter and are summarized in figs. 4.1 to 4.3.

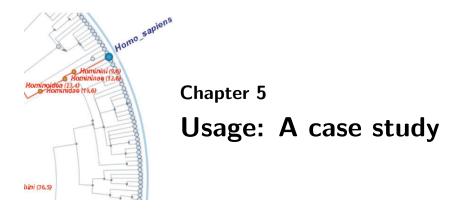| Icon	Description	Comments
	Load tree file	Opens up a file selection dialog. NWK and PhyloXML formats supported
	Capture a snapshot	Saves a TIF image in Treevolution's install folder
	Activate/Deactivate fisheye distortion on rings	F2 key has the same behavior
	Change the amount of fisheye distortion	Pops up a bar to select the interval. Click again to hide the bar
	Filter the interval of visible rings	Pops up an bar to select the interval. Click again to hide the bar
	Reset the width of sectors	
	Automatic computation of the width of sectors	Depending on the breadth and deepness of the sector
	Reset the width of rings	
F5 key	Show/Hide ring limits	
	Highlight direct/indirect ancestors and descendants	F3 key has the same behavior
	Show/Hide node labels	Node labels conveys the name and taxonomy if available
	Highlight sectors on click	
	Highlight rings on click	
	Change the locations of sectors by dragging them	F4 key has the same behavior
	Split the family of the selected node as a new sector	
	Remove the sector of the selected node	

Figure 4.1. *Radial dendrogram interaction table*

Icon	Description	Comment
	Load tree file	Opens up a file selection dialog. NWK and PhyloXML formats supported. It returns to the radial visualization.
	Capture a snapshot	Saves a TIF image in Treevolution's install folder
	Activate/Deactivate fisheye distortion on rings	F2 key has the same behavior
	Change the amount of fisheye distortion	
	Filter the interval of visible rings	Pops up an bar to select the interval. Click again to hide the bar
	Show/Hide node labels	
	Rotate node labels 45 degrees	Improves the readability when there are several horizontal lines
	Highlight columns on click	
	Reset the width of columns	

Figure 4.2. *Linear dendrogram interaction table*

Icon	Description	Comment
	Show/hide background dendrogram	The background dendrogram updates with bar selections
	Load tree file	Opens up a file selection dialog. NWK and PhyloXML formats supported. It returns to the radial visualization.
	Capture a snapshot	Saves a TIF image in Treevolution's install folder
	Sort bars on ascendent order	
	Sort bars on descendent order	
	Sort bars on alphabetical order	
	Number of nodes within each ring	Logarithmic scale
	Number of direct descendants of each node	Logarithmic scale

Figure 4.3. *Histogram interaction table*

Chapter 5

Usage: A case study

5.1 Felidae inspection

On this usage example, we will describe the steps to explore an evolutionary tree in order to find and analyze a subfamily of elements. We will explore the *Felidae* family on a tree comprising all the carnivore mammals. This tree is a subtree of the tree from [2].

After starting Treevolution (see section 1.3), the first step is to load the file *carnivore.nwk*, located at the *data* folder in Treevolution installation path. This file contains a Newick tree with information of branch lengths that convey the time until their split from the direct ancestor, in millions of years (Myrs.). To load it, we click on the *Load tree file...* button (see fig. 5.1).
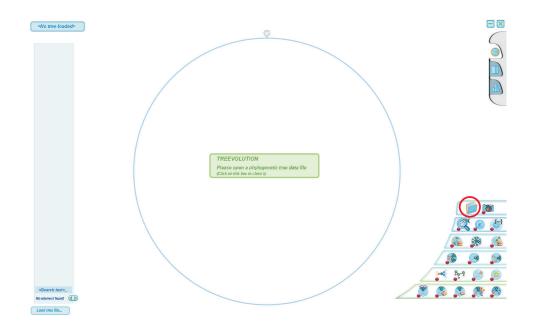
Figure 5.1. *Initial view of Treevolution. Click on the icon highlighted to open a Newick of PhyloXML tree.*

Now that the tree is loaded, the radial dendrogram displays it. Initially, we know that there are 481 nodes and we can see the whole visualization of the tree. The linear dendrogram is disabled, and the lock is open, so no node is locked (see fig. 5.2).
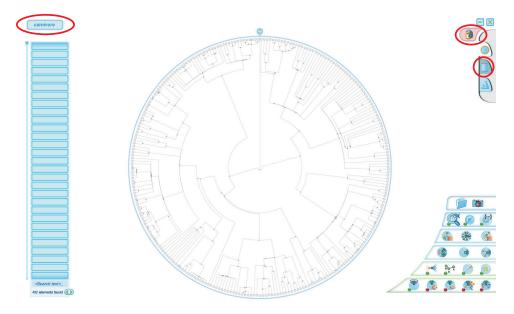

Figure 5.2. *Treevolution after loading* carnivore.nwk. *The name of the graph is displayed at the top left. At the bottom left we can see the number of elements (481). At the top right, the red lock means that there is no node locked and the linear dendrogram visualization is disabled.*

If we want to search for an specific node, the best way is to use the text search box at the left bottom. Let's write there *Felidae*. An element is found, if we click on the corresponding text box the node is selected in the dendrogram in green (see fig. 5.3). The label *Felidae (16.3)* tells us that *Felidae* appeared 16.3 Myrs. ago. The interpretation of the number depends on the meaning of the branch lengths, and always convey the sum of branch lengths from the further leaf node to the focused node. Also, a green circle passing through the node is drawn for comparison with other nodes.

By hovering the *Felidae* node, we can see its ancestor, at 35.1 Myrs. and its two direct descendants, one of them is the leaf node *Acinonyx jubartus* and the other one is hardly seen under the *Felidae* tag. We can rotate the display with the top rotation button in order to watch the labels clearer. If we click on the corresponding button or press *F3* key and hover over *Felidae* again, every ancestor and descendant, direct or indirect, are also highlighted (see fig. 5.3).
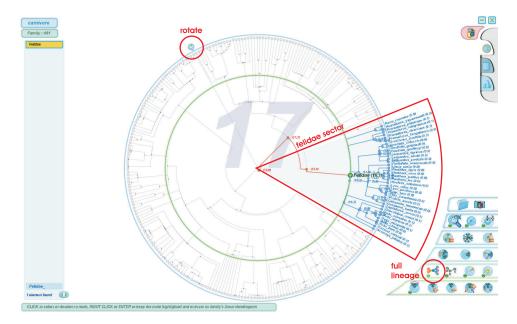

Figure 5.3. Felidae *was searched in the left text column, and the only occurrence clicked. The result is that the* Felidae *node is highlighted in green. Clicking on the full lineage node (right bottom) and then hovering over the* Felidae *highlights its family (red and blue branches), its sector (surrounded in red) and its ring (background number). Note that the node is selected but not locked. By dragging the top rotation button we can set the sector more horizontal.*

The sector dedicated to Felidae is too small. We have two options to make it bigger: to distort the radial dendrogram or to switch to a linear dendrogram of Felidae. To distort the dendrogram, we can start hiding the rings corresponding to a period of time previous to 17 Myrs. To do this, we select the proper button in the interface and an interval bar appears. By displacing the left handle up to 17 Myrs., rings from 54 to 18 Myrs are shrunk in the center of the visualization, giving more space to the other rings (see fig 5.4). Afterwards, we can create sector limits for Felidae by selecting the node and clicking the proper interface button. When a sector is split from the original 360 degrees sector, it takes half the space. This could be too much for Felidae, so by dragging the sector delimiting lines, we can reduce the breadth of the sector (see fig. 5.5). The Felidae family is clearer now, but still the close direct descendant of the *Felidae* family is hard to see. Let's distort it's ring size to solve this: click on the outer limit of ring *17* and drag it out. If you make it large enough, the node becomes more visible: it is located at the same time than *Felidae*, 16.3 Myrs. and has no special name (see fig. 5.6).
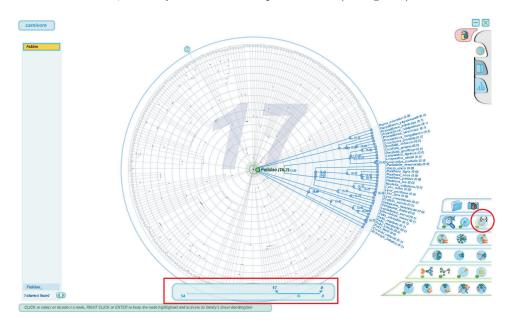

Figure 5.4. *The interval filter button at the right bottom opens an interval bar. The scrolls are set to the interval [17,0], so only nodes after 17 Myrs. ago are shown.*
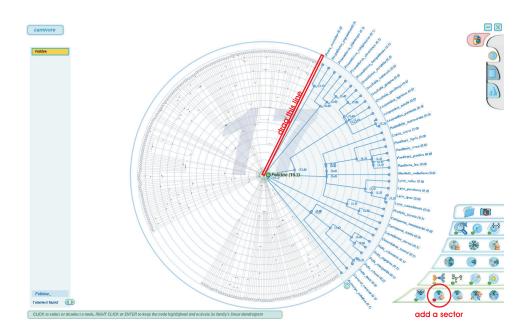
Figure 5.5. *By clicking the* add sector *button, a new sector is split for the selected node (in this case,* Felidae*). It will occupy half the dendrogram, we can drag the new sector boundary to shrink it a bit.*
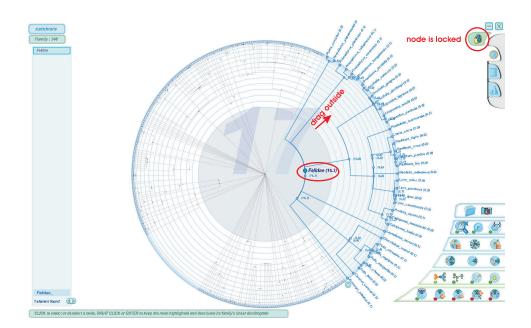
Figure 5.6. *To drag outside the 17 Myrs. ring clarifies the visualization of the nodes very close to* Felidae. *If we right click on* Felidae, *it becomes locked and its linear dendrogram is enabled.*

If we right-click the *Felidae* node, it becomes locked and the linear dendrogram
is activated. When a node is locked, to hover over other nodes does not cause any
effect, until the node is unlocked again with another right-click. By clicking on
the linear dendrogram button we switch to the linear display of the *Felidae* family.
Again, we can filter previous years (now they are columns instead of rings), show
labels and rotate them 45 degrees so they do not interfere with horizontal branches.
Two more nodes were very close to *Felidae* that weren't distinguishable on the radial
dendrogram. Now it is clear that there are two main splitting times for Felidae, 17
and 9 Myrs. You can highlight them by activating the "column highlight" button
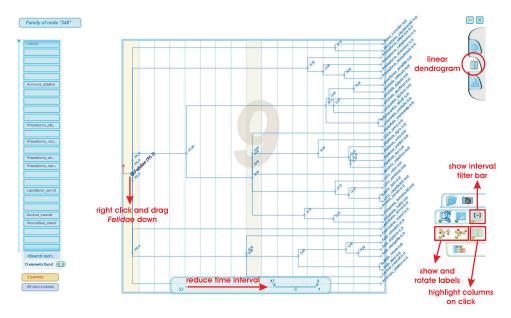and clicking on these columns (see fig. 5.7).

Figure 5.7. *The linear dendrogram has analogous interactions to those of
the radial dendrogram. In addition we can rotate labels and drag nodes in order to
clarify the visualization.*

If we go back to the radial dendrogram and right-click on a ring between 17 and 9 Myrs., the main subfamilies of *Felidae* are colored (see fig. 5.8): *Leopardus* and *Oncifelis*, *Panthera* and *Lynx*, *Felis*, and *Prionailurus*. The diversification at 9 Myrs. is related to the *Panthera* subfamily, separating it from *Lynx* and *Catopuma* families.
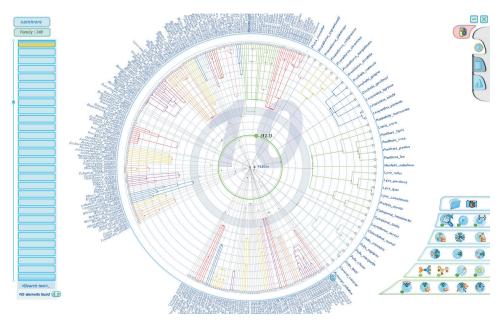

Figure 5.8. *Colored radial dendrogram after right-clicking on the ring between 17 and 9 Myrs. The Felidae family is colored in green (right).*

Histograms can give additional information about the evolution of carnivore mammals and *Felidae*, particularly the visualization of the number of nodes per ring. By clicking in the lowest visualization icon at the top right, the histogram view is shown. We select the second button in the last row of the bottom left interface, which displays the number of nodes per ring (see fig. 5.9). The histograms follow a logarithmic scale to avoid the effect of the large height of the bars where the leaf nodes are. In the illustrated histogram, we can see that the consistent branching of carnivore mammals started at the same time that the first branching of *Felidae*, about 17-18 Myrs. ago. There where some previous time frames with some activity, such as 38-36 Myrs. and 30-28 Myrs. (they have shaded bars because the time interval was filtered previously in the radial dendrogram).
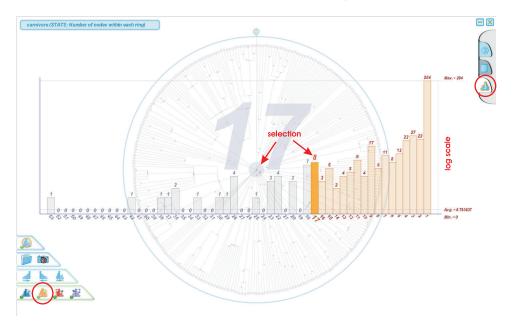

Figure 5.9. *Histogram representing the number of nodes per ring for carnivore mammals. Bars before 17 Myrs. are shaded because they were filtered on the radial dendrogram, which is displayed at the background. To hover or select a bar highlights the corresponding ring.*

Finally, we can take a snapshot of any of the three visualizations (excluding the interface buttons and search boxes). The image is stored in the installation folder of Treevolution with an automatic name that includes the type of visualization, the name of the graph and the date. The image is stored in .tif format (see fig. 5.10).
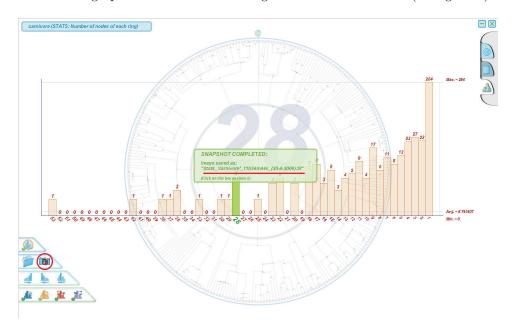

Figure 5.10. *To take a snapshot only requires to press the corresponding button. It is stored using a default path (the installation path) and name, as shown.*

Bibliography

[1] B. Fry and C. Reas, *Processing, a programming handbook for visual designers and artists*, MIT Press (2007).

[2] O. R. P. Bininda-Emonds, M. Cardillo, K. E. Jones, R. D. E. MacPhee, R. M. D. Beck, R. Grenyer, S. A. Price, R. A. Vos, J. L. Gittleman, and A. Purvis, "The delayed rise of present-day mammals," *Nature* **446**, 507–512 (2007).

[3] M. Sarkar and M. H. Brown, "Graphical fisheye views," *Communications ACM* **37**(12), 73–83 (1993).