

PoInTree: A Polar and Interactive Phylogenetic Tree

Carreras Marco, Gianti Eleonora, Sartori Luca, Plyte Simon Edward, Isacchi Antonella, and Bosotti Roberta*
Nerviano Medical Sciences srl, 20014 Nerviano (MI), Italy.

PoInTree (Polar and Interactive Tree) is an application that allows to build, visualize, and customize phylogenetic trees in a polar, interactive, and highly flexible view. It takes as input a FASTA file or multiple alignment formats. Phylogenetic tree calculation is based on a sequence distance method and utilizes the Neighbor Joining (NJ) algorithm. It also allows displaying precalculated trees of the major protein families based on Pfam classification. In PoInTree, nodes can be dynamically opened and closed and distances between genes are graphically represented. Tree root can be centered on a selected leaf. Text search mechanism, color-coding and labeling display are integrated. The visualizer can be connected to an Oracle database containing information on sequences and other biological data, helping to guide their interpretation within a given protein family across multiple species. The application is written in Borland Delphi and based on VCL Teechart Pro 6 graphical component (Steema software).

Key words: phylogenetic tree, tree visualizer, tree builder

Introduction

Thanks to the arising new technologies, in the past few years a huge amount of information has been generated on sequences, genetic maps, gene expression profiles, proteomics, and biochemical pathways. Combining all this information with evolutionary analysis in an integrated way is important in understanding gene function. For instance, proximity in the phylogenetic tree may be used to start generating hypothesis on the biological role of related genes, and in the drug discovery field it can help to identify potential cross reactivity of chemical inhibitors versus closely related targets.

In order to address this issue, we have developed a phylogenetic tree builder and visualizer, called PoInTree. PoInTree stands for Polar and Interactive Tree, as the main characteristics of the application are the visualization of trees in a polar view and its interactivity and customizability.

Several tools for visualization of small phylogenetic trees already exist, including Treeview (1) and ATV (2), and few others are available to visualize larger trees, like Hypertree (3) and Walrus (4), based on hyperbolic visualization. To meet the need of visualizing medium-large trees, without penalizing the proportional relationship among branches, we have

chosen to utilize a radial view. In our local implementation PoInTree has been interfaced with an Oracle gene-oriented database that allows retrieval of biological information related to the displayed genes.

Algorithm and Features

PoInTree takes as input a FASTA file or multiple alignment formats. Phylogenetic tree calculation is based on a sequence distance method and utilizes the Neighbor Joining (NJ) algorithm (5). It also allows displaying precalculated trees of the major protein families based on Pfam classification, once Pfam alignments are downloaded as .msf files (6).

PoInTree displays medium-large phylogenetic trees in a radial view (Figure 1). In a polar or radial view, coordinates describing each point are modulus and phase (Rho, Theta). The origin of a point is its parent (*i.e.* relative translation). The modulus represents the distance between each point and the corresponding parent and is calculated by NJ algorithm.

The space optimization algorithm finds the phase corresponding to each point. It is a recursive algorithm that starts from the tree center, moves toward and reaches any leaf, and links them all with a line. Every point resides on an arch, whose amplitude depends on the number of children of the parent, where

* **Corresponding author.**

E-mail: roberta.bosotti@nervianoms.com

This is an Open Access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

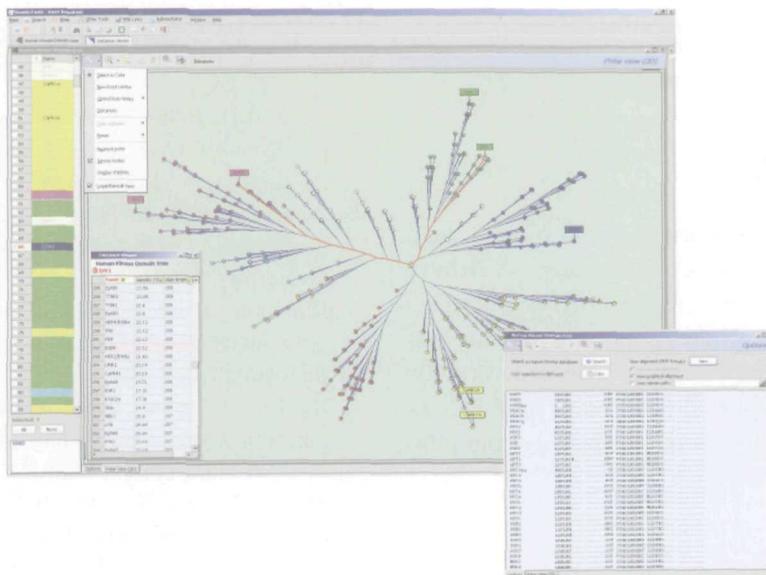


Fig. 1 PoInTree interface showing the human kinome. Branches are colored accordingly to group classification (TK in purple, TKL in light blue, AGC in red, CAMK in yellow, CMGC in green, CK1 in orange, STE in blue). Selected kinases (check box, left panel) are labeled. A red line graphically represents distance between two selected genes, and similarity value is reported in the distance table (left bottom). The alignment used to build the tree is reported in the right bottom panel.

the fraction is calculated dividing the length of the arch by the number of the children.

Theta is the media between the two angles that describe the arch. To the modulus Rho is applied a logarithm.

Features

Searching

Genes represented on the tree can be searched by key words or selected from a gene list. The corresponding labels will be interactively highlighted in the tree. Multiple selections are available. Checked sequences can be exported in FASTA format or sent to search engines to retrieve additional information.

Color-coding

Each leaf is represented by a pellet and a label. Labels can be hidden. Single leaf or leaves belonging to a node (children) can be simultaneously selected and colored.

Tree center

The new tree center function allows rebuilding the tree starting from a different center. This function also optimizes the tree in the space, allowing a better visualization and printing of the tree.

Open and close nodes

Nodes can be opened and closed. The function does not act on distances, but only on branches visualization. Closing a node will mask all the children associated with that node.

Distances

PoInTree allows, interactively, the visualization and calculation of distances between two leaves. Once a leaf is selected, a table is created with percent identity and alignment length of all the leaves versus the selected one. Following mouse movement, a red line is drawn between two points. The calculation of the line is based on an iterative algorithm made by two nested cycles that start from both points and go up until the intersection between the two ways is reached, that can be in extreme case the center of the tree. Once the intersection point is found, two red lines are drawn using the same iterative algorithm.

Hardware requirements and software availability

The application is written in Borland Delphi and based on VCL Teechart Pro 6 graphical component (Steema software). It currently runs on Microsoft Windows NT, 2000, and XP. The PoInTree can be

accessed at <http://geneproject.altervista.org/> and is available upon request.

Conclusion

Colored phylogenetic trees are essential tools to help identify the relationships between genes. We have presented PoInTree, a new user-friendly visualization program for representing phylogenetic trees in a customizable and graphical way. After customization the tree pictures can be exported as bitmaps or Windows Metafiles (wmf, emf) or simply copied to the clipboard.

References

1. Page, R.D. 1996. TreeView: an application to display phylogenetic trees on personal computers. *Comput. Appl. Biosci.* 12: 357-358.
2. Zmasek, C.M. and Eddy, S.R. 2001. ATV: display and manipulation of annotated phylogenetic trees. *Bioinformatics* 17: 383-384.
3. Bingham, J. and Sudarsanam, S. 2000. Visualizing large hierarchical clusters in hyperbolic space. *Bioinformatics* 7: 660-661.
4. Hughes, T., *et al.* 2004. Visualising very large phylogenetic trees in three dimensional hyperbolic space. *BMC Bioinformatics* 5: 48.
5. Saitou, N. and Nei, M. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* 4: 406-425.
6. Bateman, A., *et al.* 2004. The Pfam protein families database. *Nucleic Acids Res.* 32: D138-141.