

TreeSnatcher Plus

Reading phylogenetic trees from pixel images

Thomas Laubach and Martin J. Lercher

Bioinformatics group, Heinrich-Heine-University, Düsseldorf, Germany*

Introduction

TreeSnatcher Plus automates the generation of a machine-readable representation of multifurcating phylogenetic trees contained in pixel images. It uses image processing methods to prepare the image and to detect the tree structure. The user supervises the semi-automatic recognition process and makes corrections to the image and to the topology where necessary. At the end the program produces a Newick tree code that represents the tree structure optionally including branch lengths.

Workflow

Load an image that shows a phylogenetic tree (cf. fig. 1). Keep the parts of the foreground that represent the tree.

Sharpen, lighten, invert, despeckle etc. the whole or parts of the image in order to turn the foreground dark and the background light (see figs. 6 and 7).

Use pencil and rubber to modify the image and to fill gaps in lines. The tree structure and other structures, i. e. characters, must not overlap.

Let the program separate the foreground from the background using binarization and thinning. Signal the program which part of the foreground is the tree. Use floodfilling for this purpose (figs. 2, 6 and 7).

The application then determines the locations of inner nodes and tips. Add or remove nodes as needed.

Use the branch detection function. You may delete branches or add branches by dragging connections between two nodes.

Repeat some of these steps if you need to modify the tree topology.

Assign species names and specify branch lengths (cf. fig. 3). You can also scale the tree using a scale bar that is shown in the image.

Select a node as reference node for the recursive construction of the Newick tree code.

Let the program build the Newick tree code as shown in fig. 4. Save it to the clipboard, or export it into a text file.

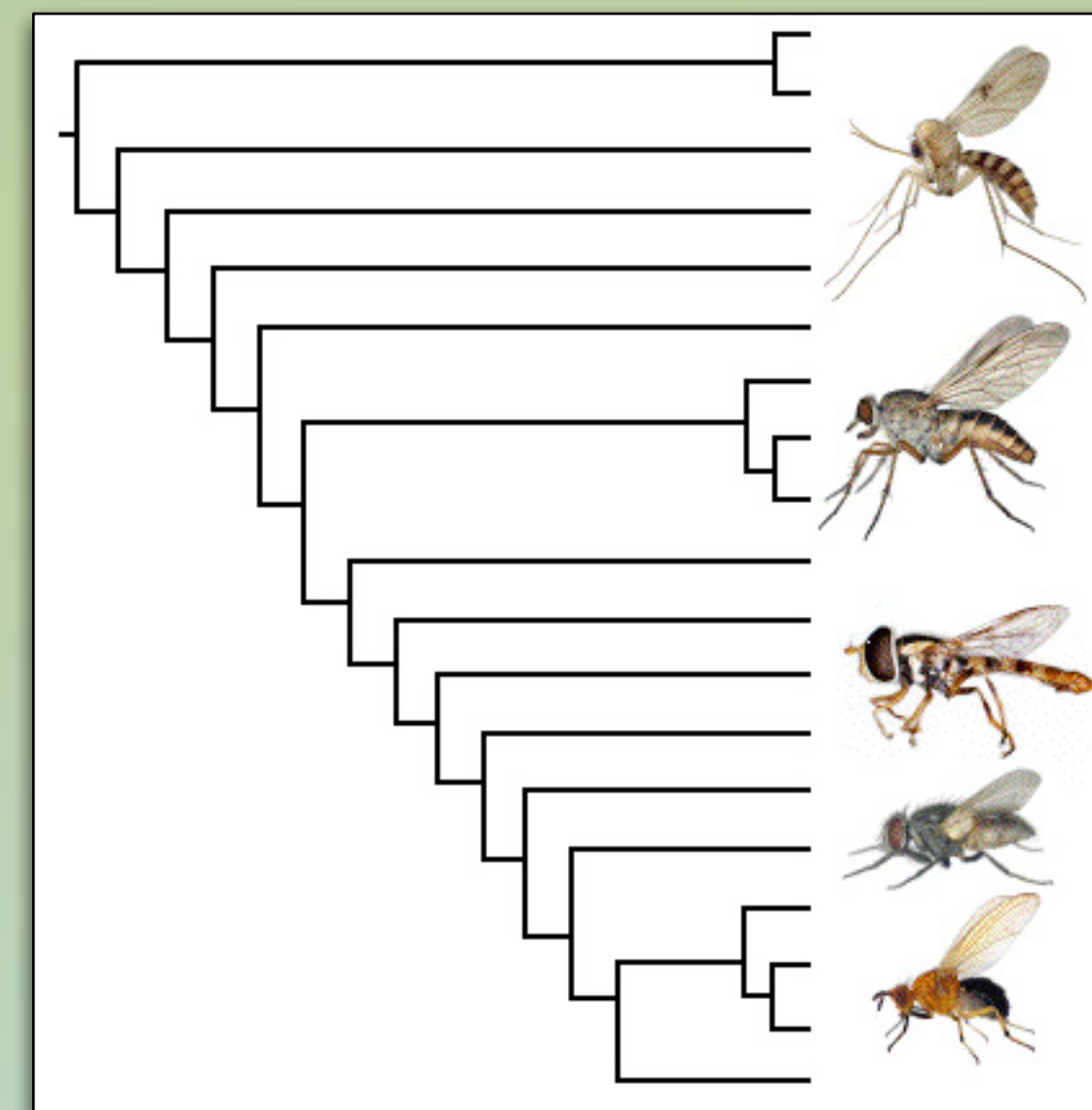


Figure 1: Source image "Flytree" used with kind permission from Brian Wiegmann [4]

- 1 Trim and cut image
- 2 Prepare foreground
- 3 Mark tree

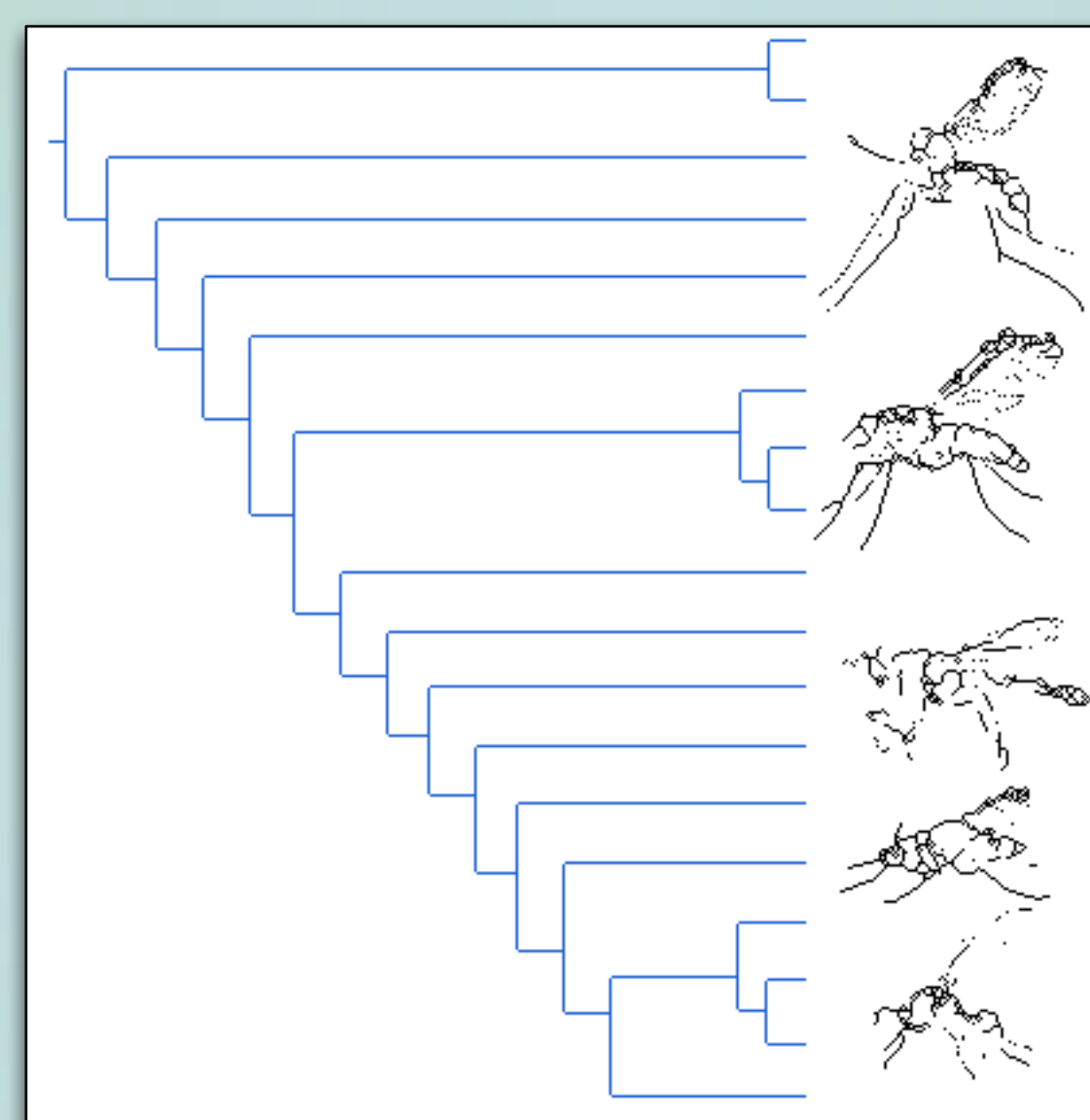


Figure 2: The foreground that represents the tree is colored in blue

- 4 Determine nodes
- 5 Determine branches
- 6 Modify topology
- 7 Scale tree

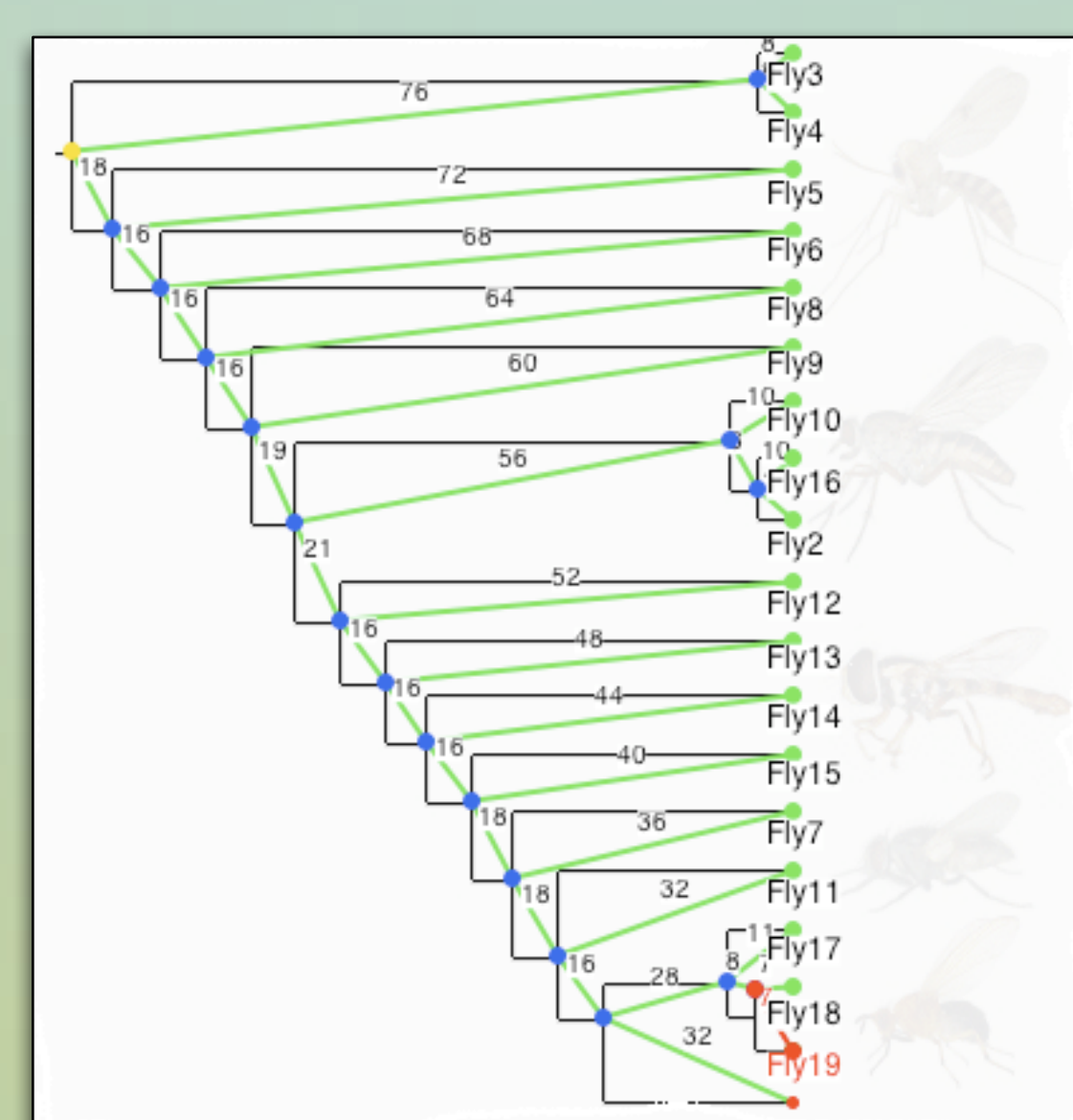


Figure 3: Identified topology with nodes, branch lengths and taxon names

Your Phylogenetics application

- 9 Export Newick tree code

```
((Fly4:72, (Fly5:68, (Fly6:64, (Fly8:60, ((Fly2:52, (Fly12:48, (Fly13:44, (Fly14:40, (Fly15:36, (Fly7:32, (Fly19:32, (Fly11:11, (Fly17:7, (Fly18:7):8):28):16):18):16):16):21, (Fly9:10, (Fly10:10, (Fly16:10):8):56):19):16):16):18, (Fly1:8, (Fly3:8):76);
```

Figure 4: The Newick expression for the tree depicted in the source image (see fig. 1)

- 8 Determine Newick tree code

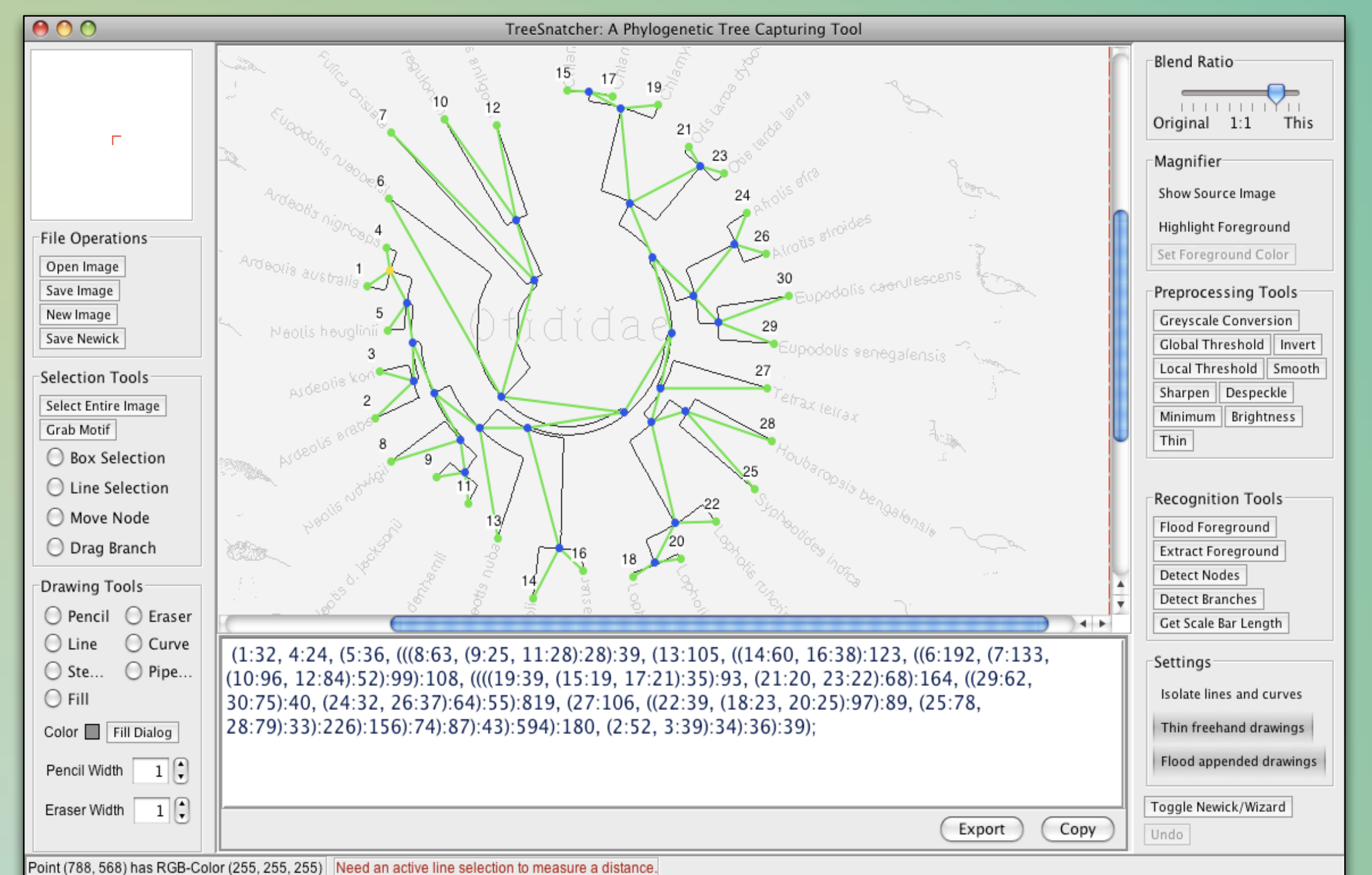


Figure 5: TreeSnatcher Plus main screen displaying the Otididae phylogeny from Pietra et al. [2]

Program features

- Java application for Windows, Linux and MacOS X
- Graphical User Interface (GUI)
- Convert an image of a tree into a tree
- Save and restore the program state ("Checkpointing")
- Undo functionality for all operations
- Accepts the formats JPG, TIF and PNG
- Preprocess the entire image or image portions
- Scale the tree using a scale bar that is shown in the image
- Type in branch lengths, or let the program calculate them
- Program ensures that the tree topology is valid
- Export a Newick expression into a text file
- Save the current processing state as an image file

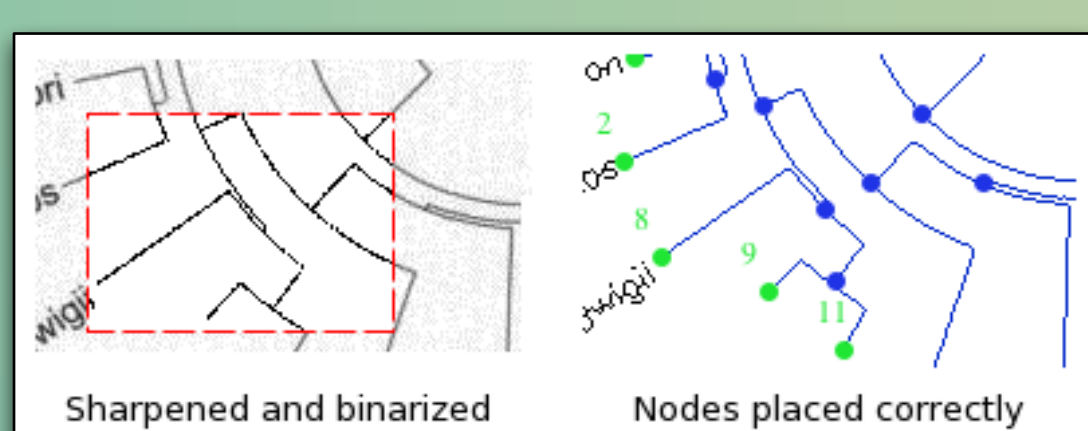


Figure 6: A portion of the Otididae image [2] after adept preprocessing and resulting nodes placement

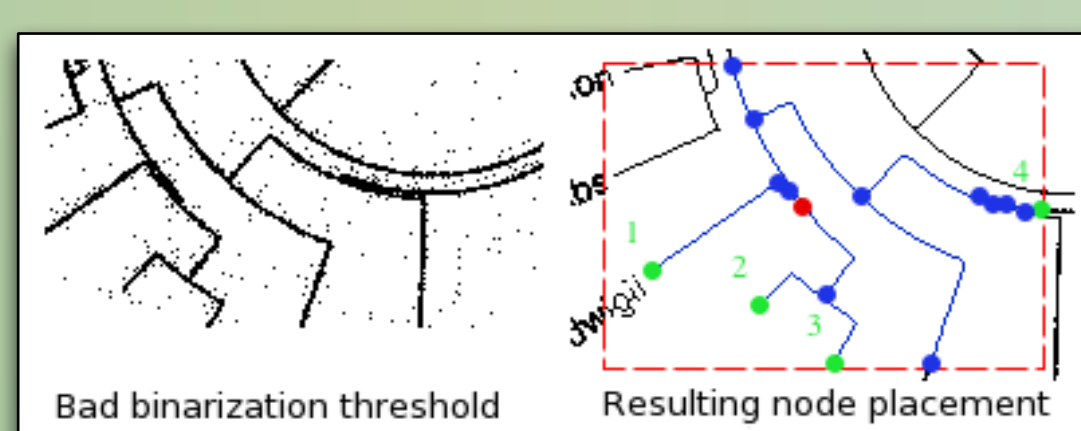


Figure 7: The same portion of the Otididae image [2] after inapt preprocessing

References

1. Laubach, T.; Haeseler, A. von: TreeSnatcher: Coding trees from images. In: Bioinformatics 23 No. 24 (2007)
2. Pietra, C. ; Lieckfeldt, D. ; Frahnert, S ; Fickel, J.: Phylogenetic Relationships and Ancestral Areas of the Bustards (Gruiformes:Otididae), Inferred from Mitochondrial DNA and Nuclear Intron Sequences. In: Molecular Phylogenetics and Evolution 23 (2002)
3. Rambaut, A. (2000) TreeThief: a tool for manual phylogenetic tree entry. <http://evolve.zoo.ox.ac.uk/software/TreeThief/main.html>
4. Wiegmann, Brian M.: FLYTREE: NSF Assembling the Tree of Life EF-0334948, Building the Dipteran Tree of Life: Cooperative Research in Phylogenetics and Bioinformatics of True Flies (Insecta: Diptera). <http://www.inhs.uiuc.edu/cee/FLYTREE/>, 2004

Conclusion

We presented *TreeSnatcher Plus*, an extended and re-conceptualized version of TreeSnatcher [1], that converts an image of a phylogenetic tree into a Newick format expression with user assistance. It is well fitted to support the research on complex phylogenies and will be helpful if one wants to combine phylogenies based on a collection of already published trees that are not electronically available. Moreover, *TreeSnatcher Plus* is also suited for educational purposes as it allows the user to construct his own trees and to modify existing topologies. It is the only software for the task except for *TreeThief* [3] for classic Macs.