Png tree converter

ADP Final Project

Aleksandra Cupriak, Agnieszka Kowalewska, Joanna Krawczyk

Project motivation

Struggle to analyse, convert and download phylogenetic trees found in scientific articles.

General idea

- Build a browser tool for converting phylogenetic trees in .png image format to newick format.
- Use custom line and node detection method (binary array-based).
- Create a graph on nodes’ relative spatial location.

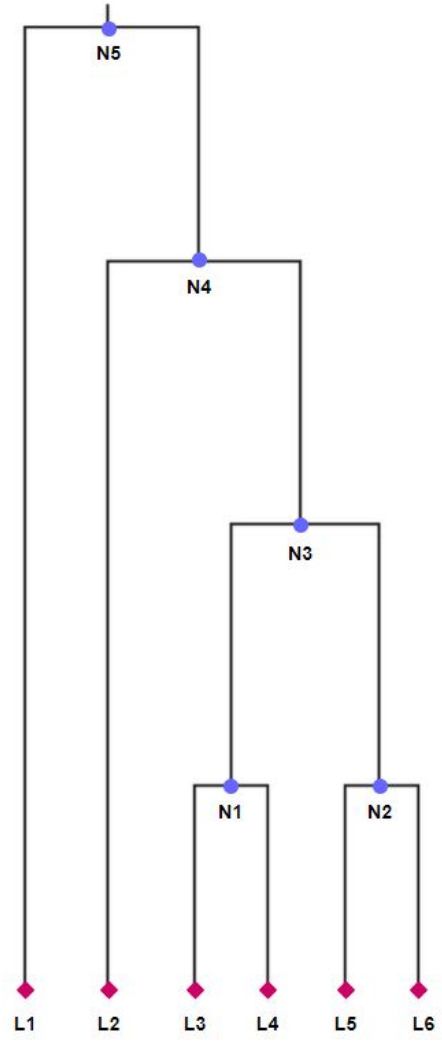
What defines a particular tree?

<u>Location of internal nodes and leaves.</u>

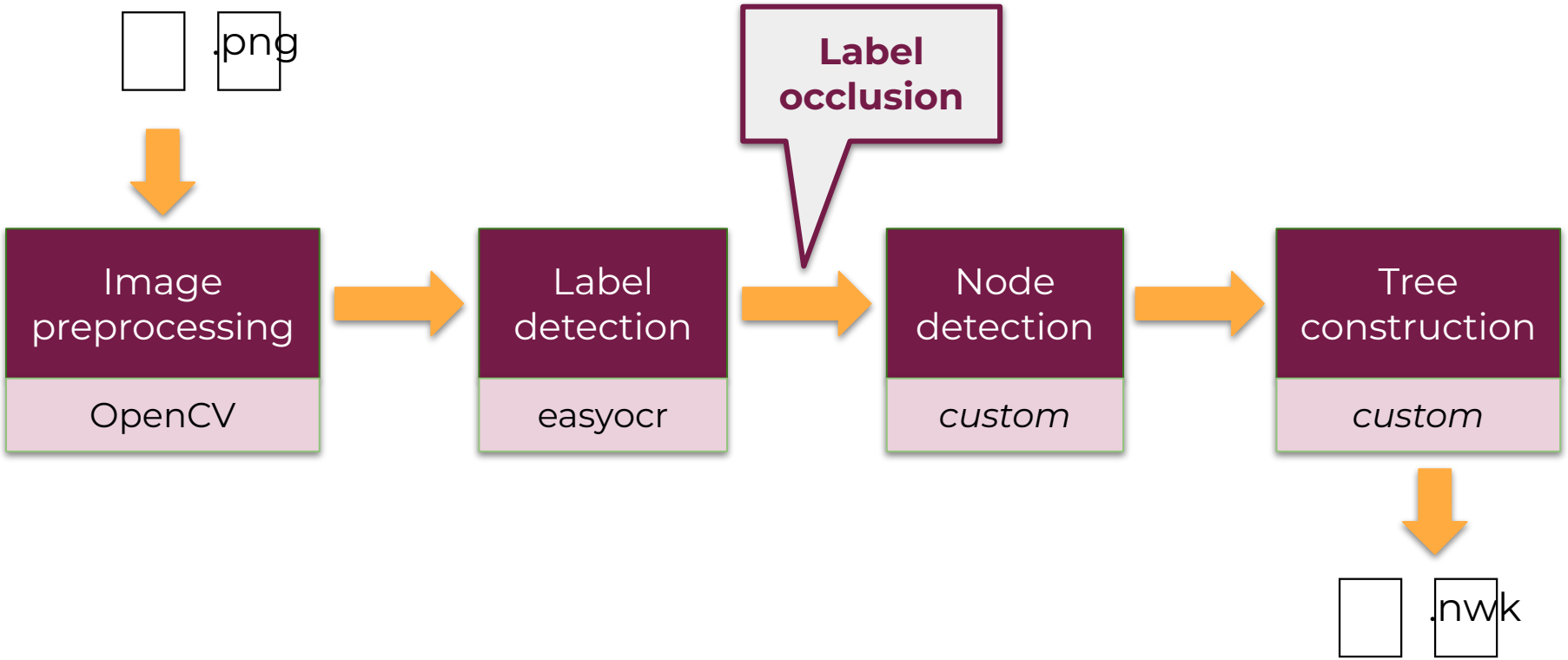
How to construct a graph based on that?

1. Sort internal nodes by y-coordinate.
2. For each subsequent node:
    - find the nearest left and right leaves,
    - remove found leaves,
    - create new leaf.
3. Stop when less than 3 leaves remain and create an artificial root.

Generate newick string based on constructed graph starting from the root - recursively.

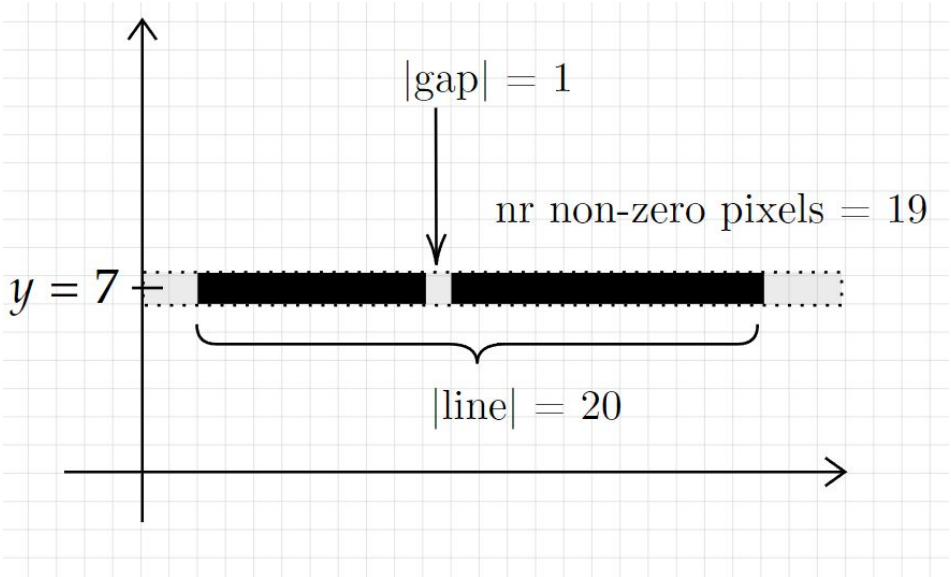


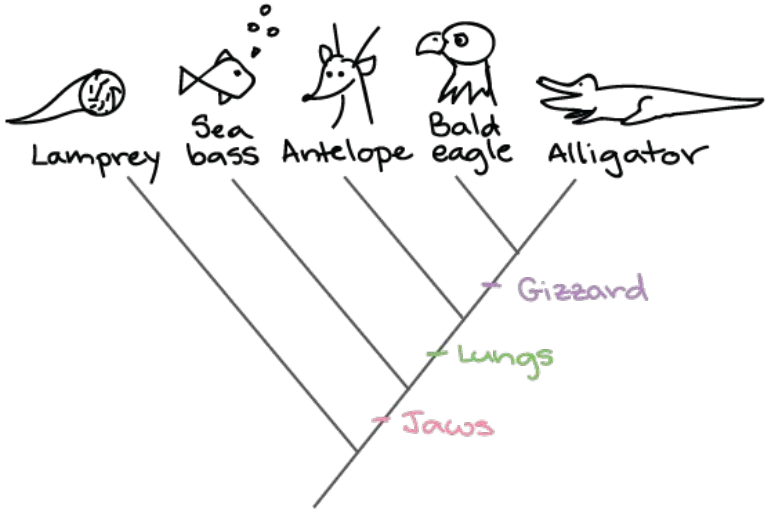
Pipeline


.png
Label occlusion
Image preprocessing
OpenCV
Label detection
easyocr
Node detection
custom
Tree construction
custom
.nwk

Line detection

We assume all lines are nearly vertical or horizontal.
We operate on binary matrix representing the image.

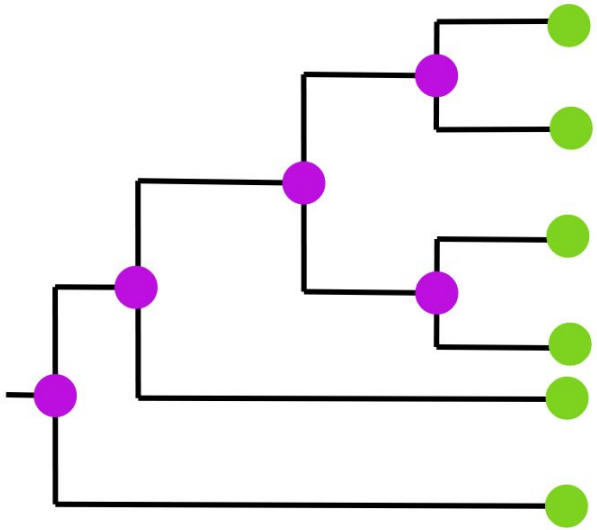




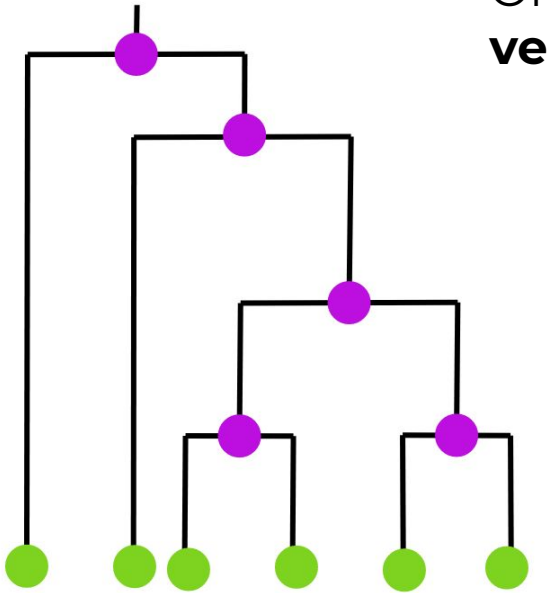
https://www.khanacademy.org/science/ap-biology/natural-selection/phylogeny/a/building-an-evolutionary-tree

Tree orientations

Orientation: **horizontal**

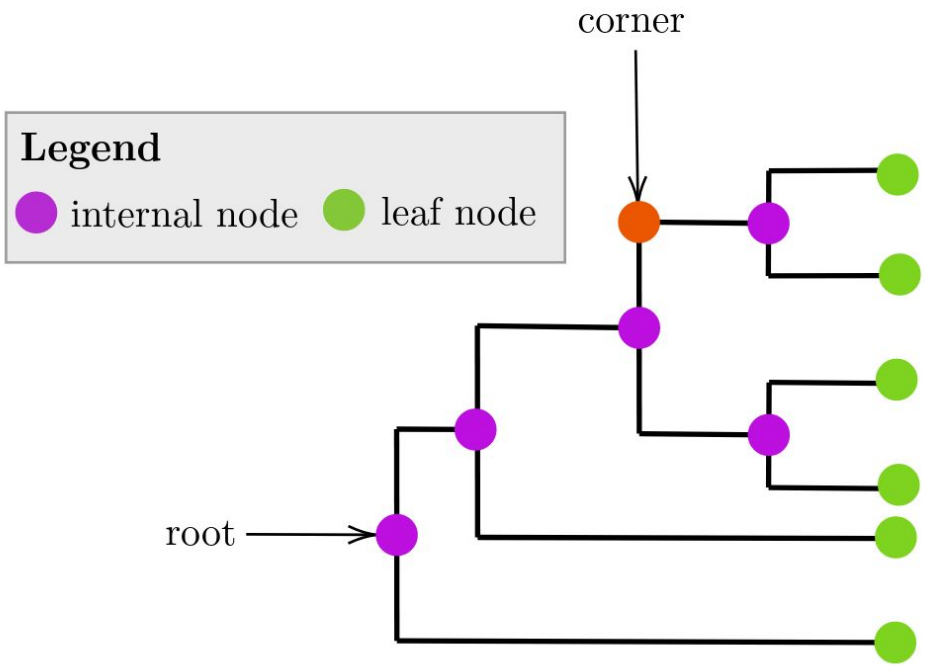


Orientation: **vertical**



Node detection

1. Find all intersections (without **corners**) between elongated horizontal and vertical lines.
2. Replace spatially close intersections with one having mean coordinates.
3. Depending on the orientation, find leaves.
4. Filter leaves (similar to p. 2).



Requirements and limitations

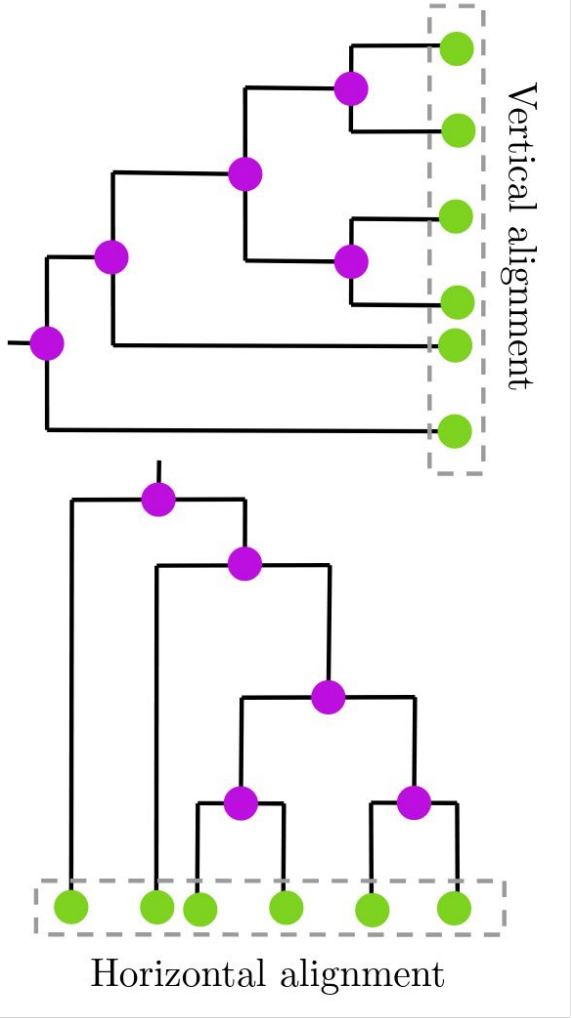
Requirements concerning tree type:

- rooted, binary,
- oriented vertically or horizontally,
- with all lines nearly horizontal or vertical,
- leaf alignment requirements.

Limitations:

- not applicable to trees with textual data other than leaf labels,
- no account for branch lengths.

❗ Streamlit app runs locally - not on Streamlit Community Cloud (problems with OpenCV).



Case study

Streamlit

Png Tree Converter

This tool was created as a final project for ADP course at the University of Warsaw. It is designed to help analyse phylogenetic trees from publications that do not include trees in textual format such as .newick or .phylo. Here, we introduce a **Png Tree Converter** as an online tool for converting trees in .png formatinto a selected textual one. Try it yourself, upload a snapshot of a tree!

Choose a file with a tree image

Drag and drop file here

Limit 200MB per file

Browse files

Upload a file

Choose parameters

Orientation

horizontal

Minimum frequency

100

Intersection threshold

5

choose parameters

Resize factor

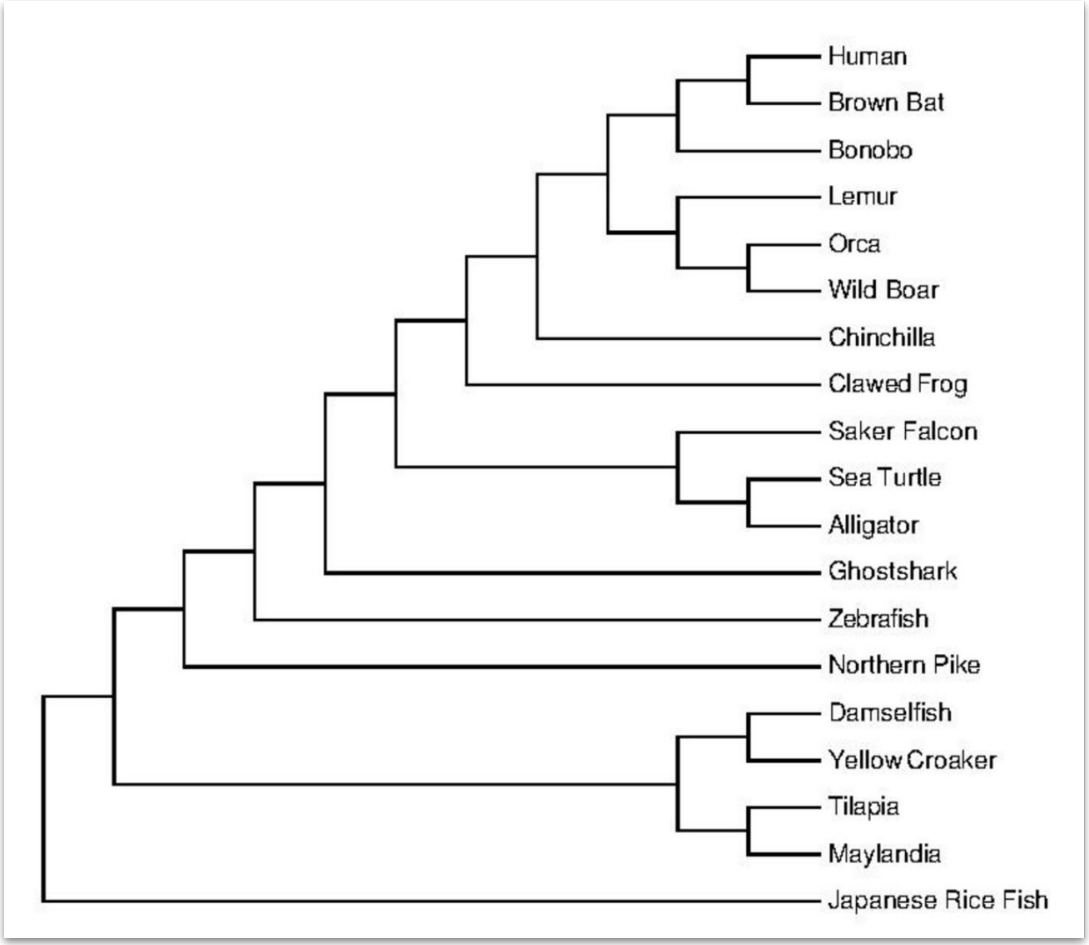
0.25

0.00

5.00

Step 1

Original image of phylogenetic tree

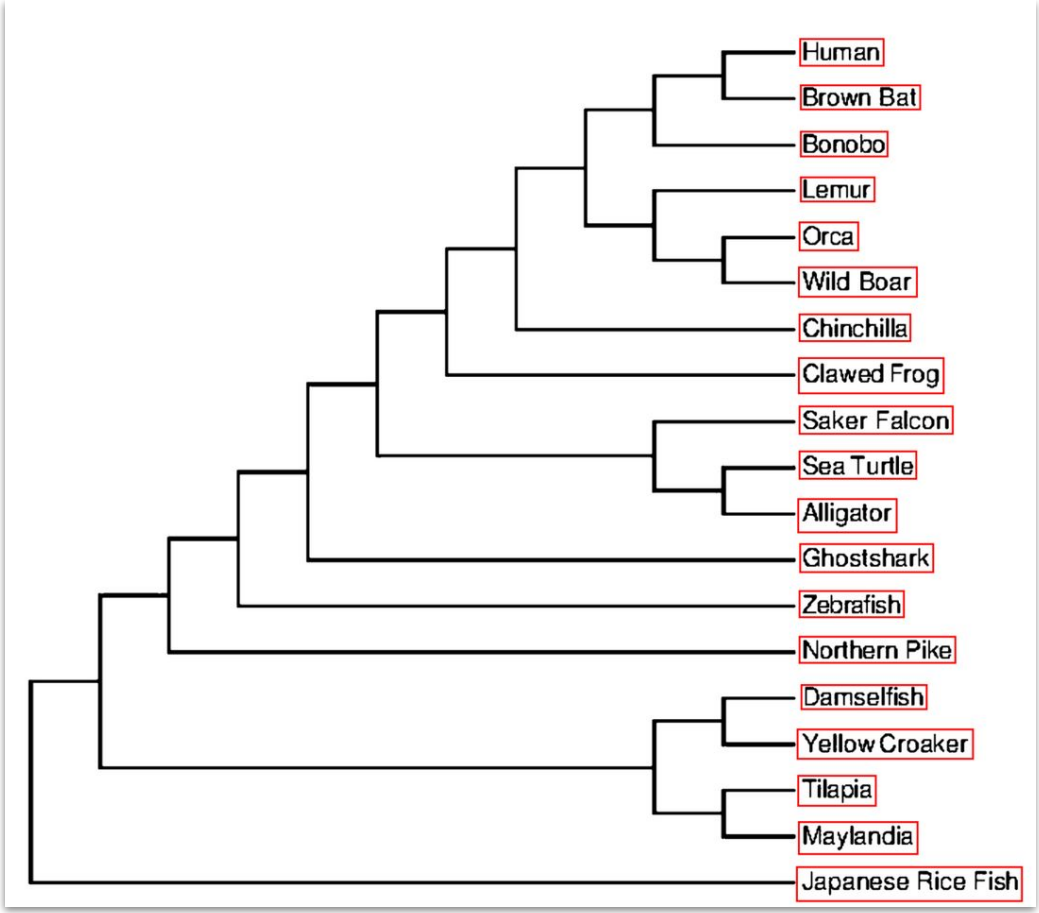


https://commons.wikimedia.org/wiki/File:Phylogenetic_Tree.pdf

Step 2
Label detection

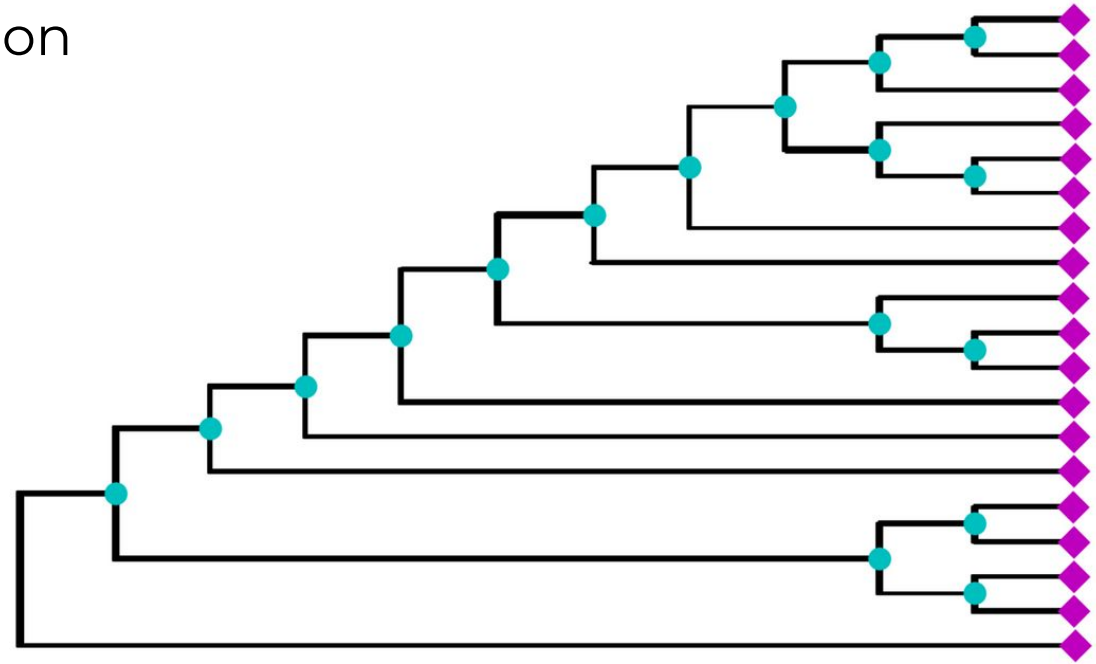
Detected labels:

Human, Brown Bat, Bonobo, Lemur, Orca, Wild Boar, Chinchilla, Clawed Frog, Saker Falcon, Sea Turtle, Alligator, Ghostshark, Zebrafish, Northern Pike, Damselfish, Yellow Croaker, Tilapia, Maylandia, Japanese Rice Fish



Step 3

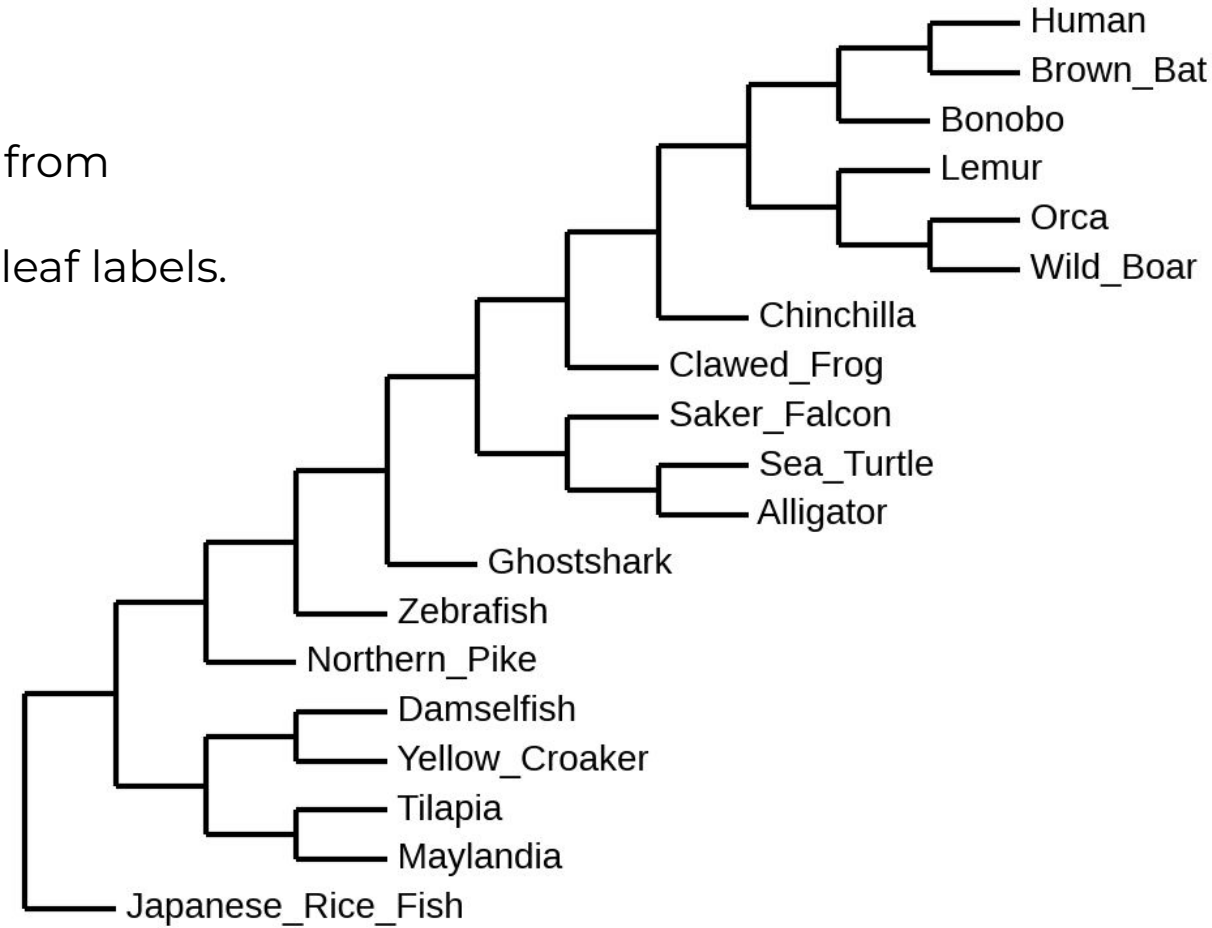
Internal nodes and leaves detection



Step 4

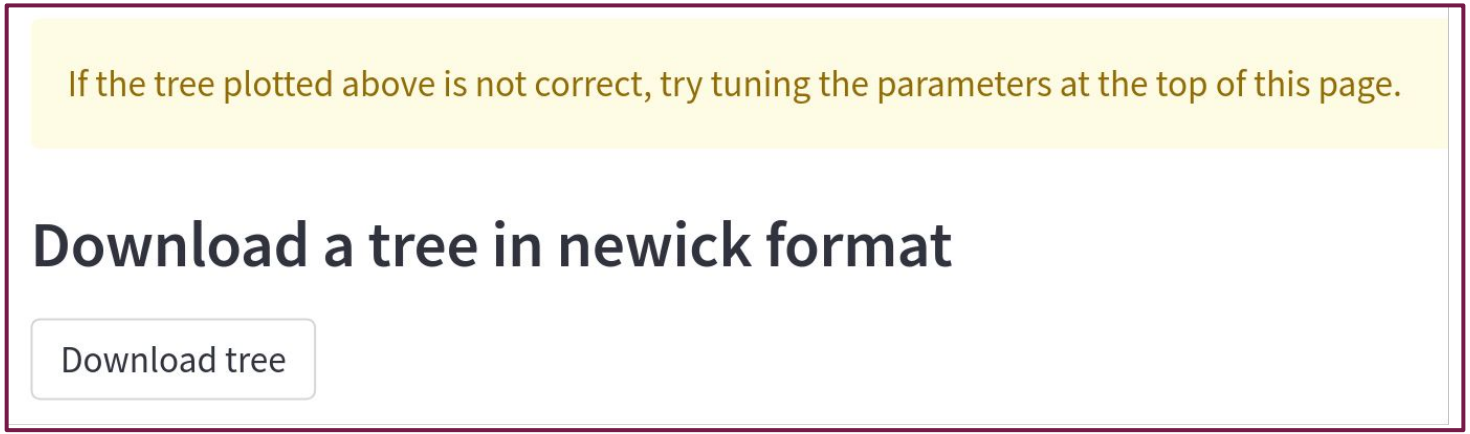
Result tree

- Infer tree structure from nodes' locations.
- Combine tree with leaf labels.



Step 5

Newick file download




```
1 ((((((((((Human,Brown_Bat),Bonobo),(Lemur,(Orca,Wild_Boar))),Chinchilla),Clawed_Frog),-
  (Saker_Falcon,(Sea_Turtle,Alligator))),Ghostshark),Zebrafish),Northern_Pike),-
  ((Damselfish,Yellow_Croaker),(Tilapia,Maylandia))),Japanese_Rice_Fish);
```

Thank you
for your attention!