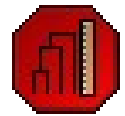
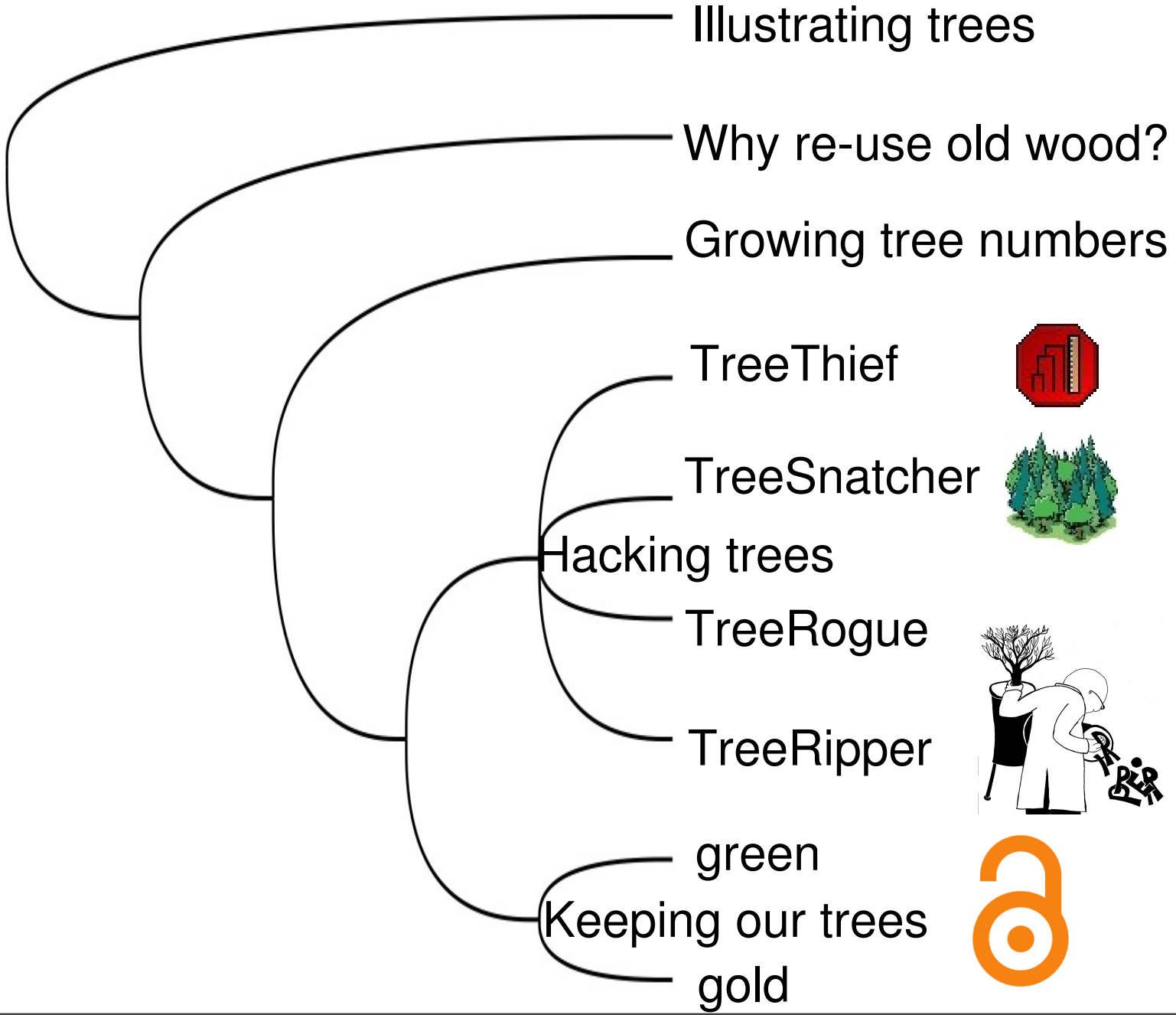


Hacking the JPEG/PDF tree format

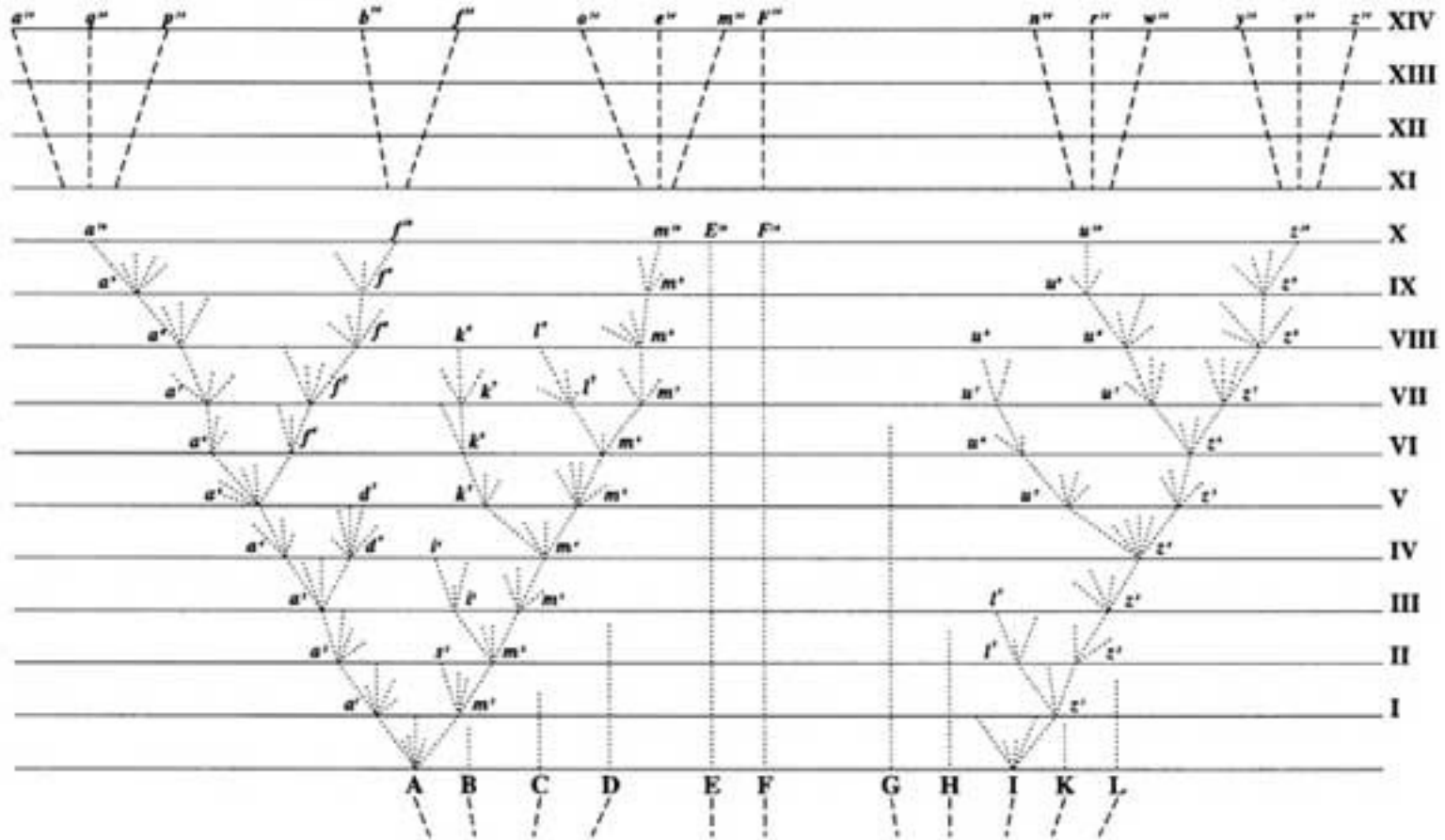
Joseph Hughes



University
of Glasgow

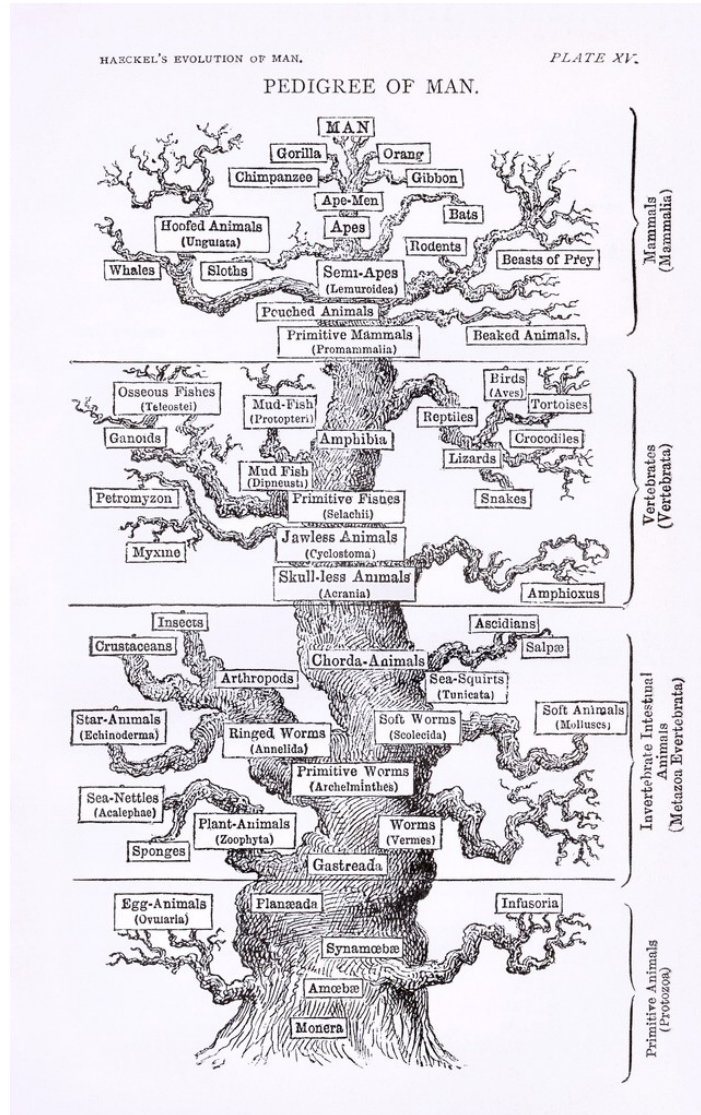


In the past



Darwin, C. R. The origin of species by means of natural selection. 1859

Weird and wonderful trees!



Why re-use old wood?

- To evaluate comparative data (evolution, ecology, biogeography, disease)
- To use as inputs for building larger trees (constraints, supertrees, megatrees)
- To study the effects of methodology (priors on tree shape)

Aggregate data in useful resources

Why re-use old wood?



TIMETREE is a public resource for knowledge on the timescale and evolutionary history of life. Search the database below or go to the TIMETREE OF LIFE for other resources.

About Search Book Resources News FAQs Contact

TimeTree Search

Find Time of Divergence

(Example: Homo sapiens, Lagomorpha, dog)

Taxon A:

Taxon B:

Clear Search

Search by Author

Last Name:

Clear Search

Citing TimeTree:

Kumar S & Hedges SB (2011)
TimeTree2: species divergence times on the iPhone. *Bioinformatics* 27:2023-2024

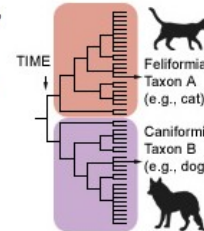
[\[Download PDF\]](#)

How It Works

How It Works

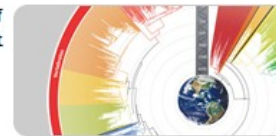
Two species or higher taxa are queried (e.g., cat and dog).

TimeTree compares all taxa in one inclusive group (e.g., Feliformia) with those in the other group (e.g., Caniformia) to find all published times of divergence for the evolutionary split.



Poster

The TimeTree of Life from an assembly of 1610 individual timetrees. Order an offset print or download the file to print your own.



TimeTree of Life Book



Download chapters for FREE of this authoritative reference work which brings together, for the first time, experts on all major groups of organisms to assemble a timetree of life.

Android App Now Available!



TimeTree for the Android is now available. Visit the Android Market to download the latest version of our TimeTree App so you can compare species on the go!

iPhone App

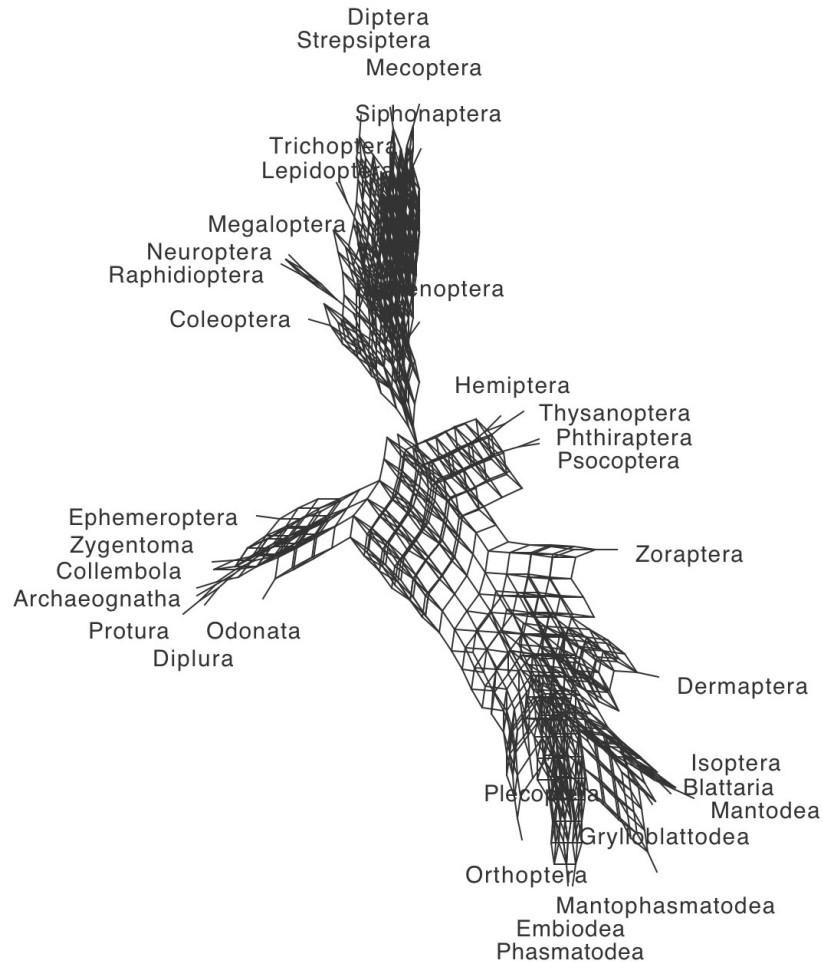


Download our FREE mobile app to compare species on your iPhone. An update will be available next week which includes our new data.

Coming Soon



Guide researchers



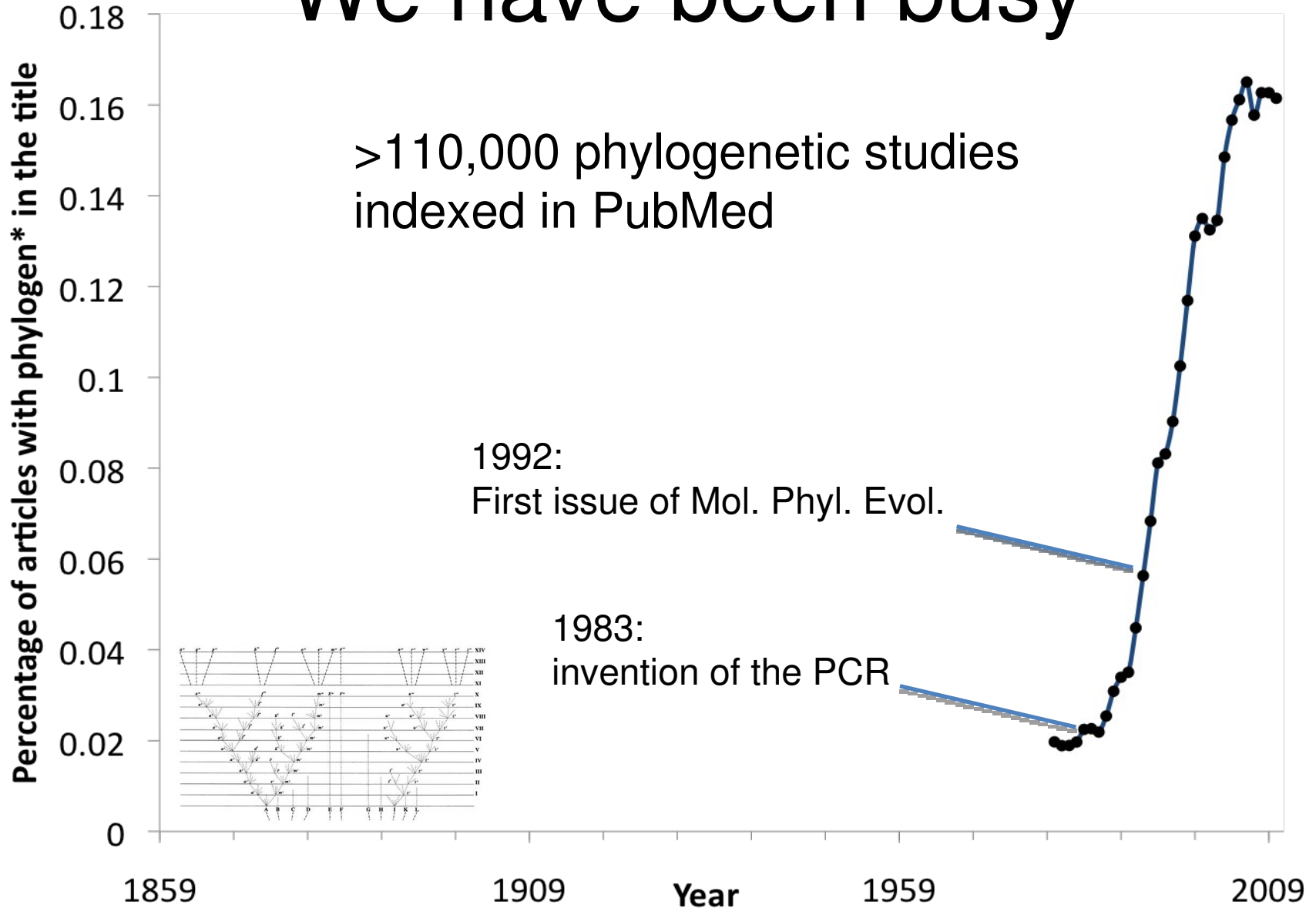
And

- To determine how phylogenetic hypotheses have changed over time
- Because the tree cannot be readily replicated (morphological characters, too time consuming, too expensive for the taxpayer)

We need to re-use!

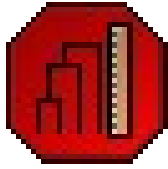
We need to be more green!

We have been busy



Different approaches





TreeThief

- A tool for manual phylogenetic entry by Andrew Rambaut
- MacOS 9
- <http://goo.gl/BFM2N>

TreeRogue

- TreeThief-like approach by Nick Matzke
- GraphClick (\$8) and R script
- <http://goo.gl/eunO2>



TreeSnatcher and +

- A semi-automated approach by Thomas Laubach
- GUI-driven Java app
- Multifurcating trees in any shape
- <http://goo.gl/Das63>

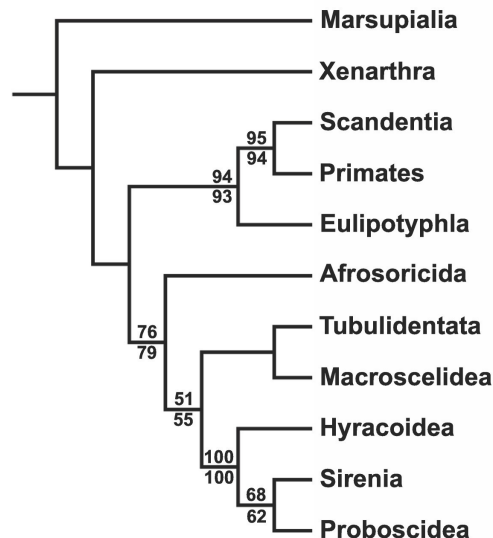
Hacking the JPEG/PDF format

This is a lot of 12 point text to test the ocr code and see if it works on all types of file format.

The quick brown dog jumped over the lazy fox. The quick brown dog jumped over the lazy fox. The quick brown dog jumped over the lazy fox. The quick brown dog jumped over the lazy fox.

This is a lot of 12 point text to test the ocr code and see if it works on all types of file format.

The quick brown dog jumped over the lazy fox. The quick brown dog jumped over the lazy fox. The quick brown dog jumped over the lazy fox. The quick brown dog jumped over the lazy fox. The quick brown dog jumped over the lazy fox.



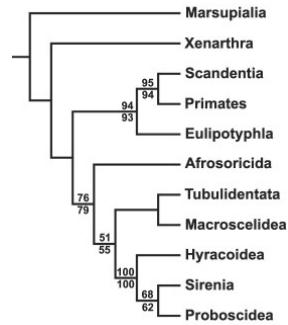
```

('Marsupialia',('Xenarthra',
(('Eulipotyphla',
('Scandentia','Primates')),
('Afrosoricida',
(('Tubulidentata','Macroscelidea'),('Hyracoidea',
('Sirenia','Proboscidea')))))));
    
```

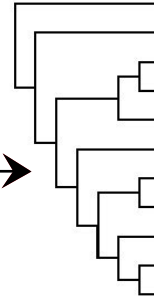
TreeRipper



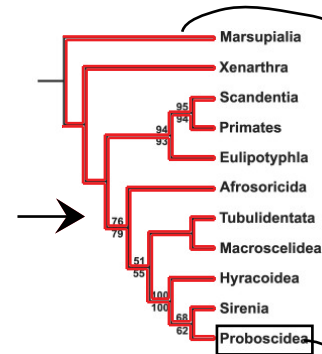
Input image



Cleaned image



Contour detection



```
# NEXUS
Begin TREES;
```

Translate

- 1 'Marsupialia',
- 2 'Xenarthra',
- 3 'Scandentia',
- 4 'Primates',
- 5 'Eulipotyphla',
- 6 'Afrosoricida',
- 7 'Tubulidentata',
- 8 'Macroscelidea',
- 9 'Hyracoidea',
- 10 'Sirenia',
- 11 'Proboscidea'

```
;
TREE TREE RIPPER = (1,(2,(((3,4),5),(6,((7,8),(9,(10,11))))))));
```

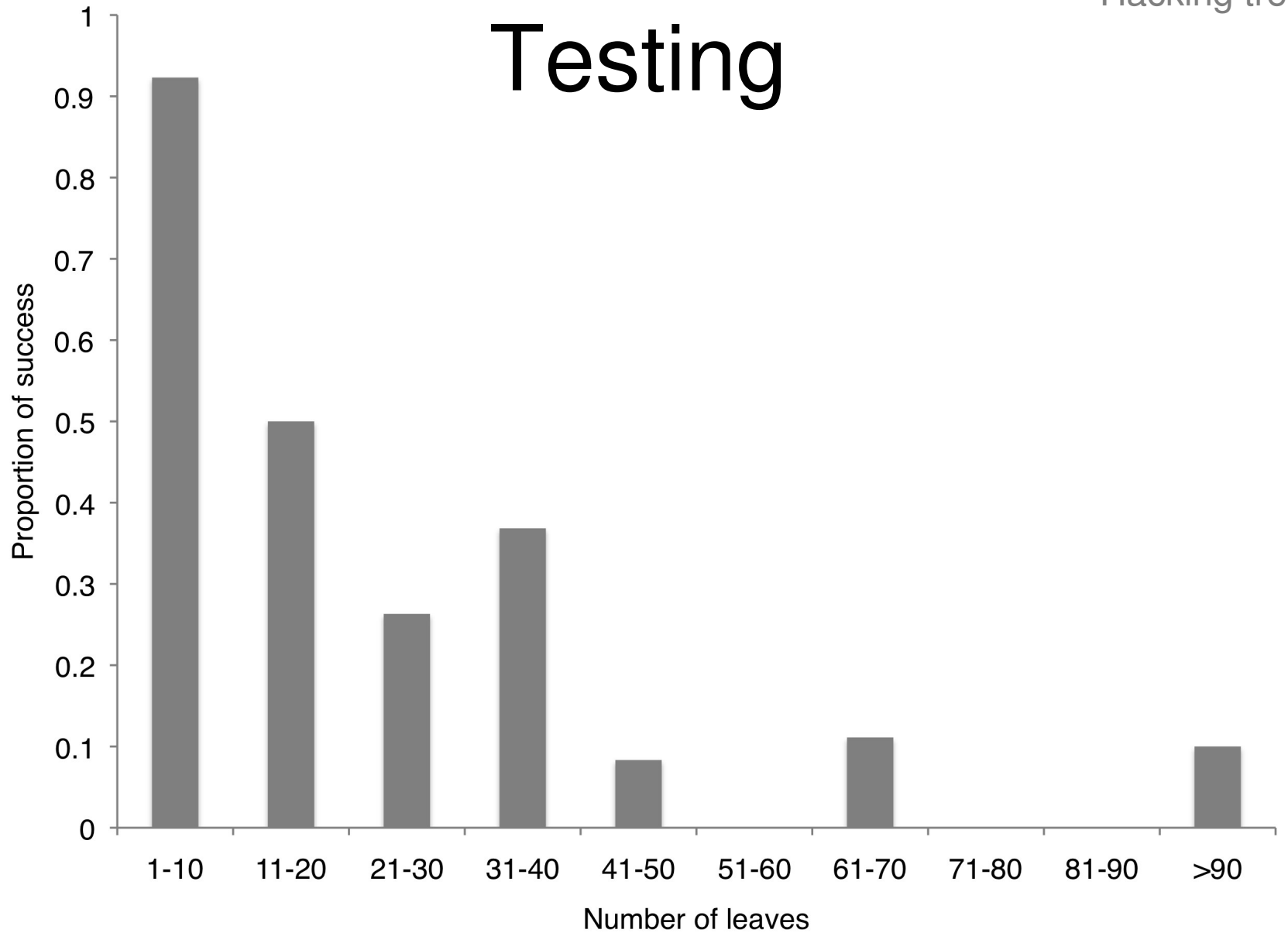
End;

OCR of labels

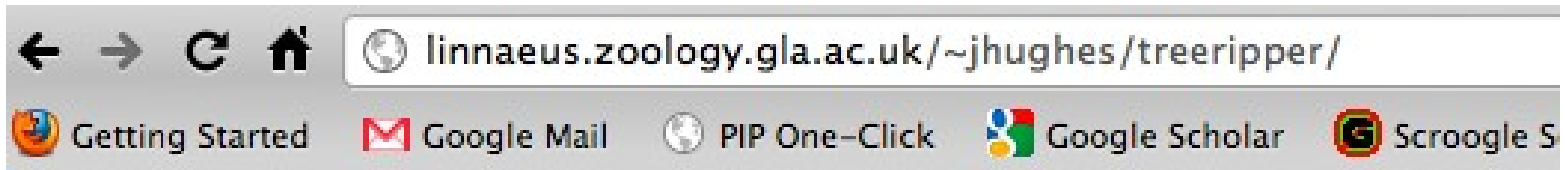
tesseract-ocr

Conversion to text format

Testing



http://goo.gl/EZ67K



Upload tree image

Name:

Email:

Choose a tree image (jpg/jpeg, png or gif) to upload:

Choose File

No file chosen

Submit

Short URL: goo.gl/EZ67K

Keeping our trees (random thoughts)

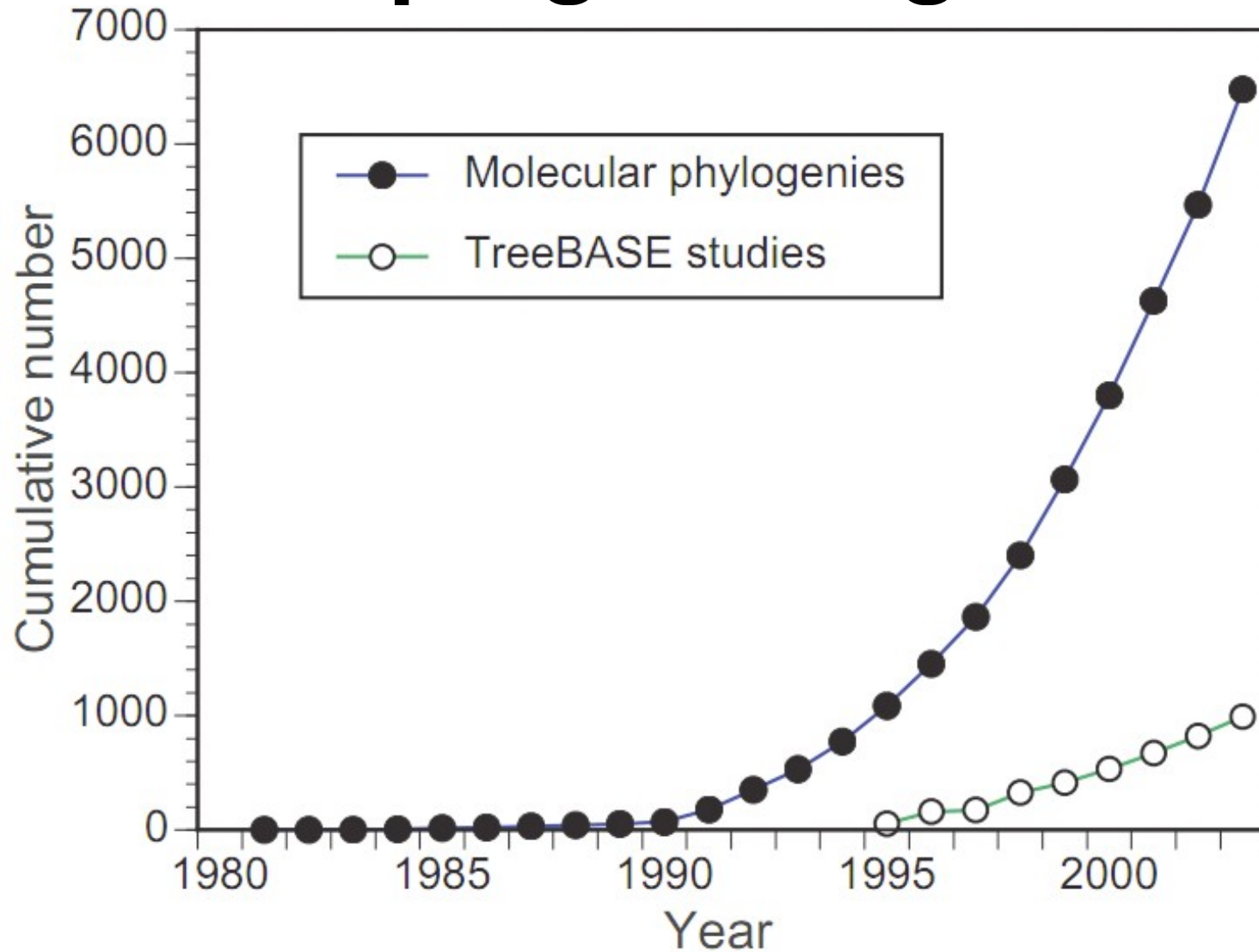
Keeping our trees

- **Green OA Self Archiving**



- **Gold OA Publishing**

Keeping trees green



Nature Precedings : doi:10.1038/npre.2011.6556.1 : Posted 26 Oct 2011

rees

Turning

Acrobat File Edit View Window Help

Hughes - 2011 - TreeRipper web application towards a fully automated optical tree recognition

Create [Icons]

1 / 4 [Navigation]

Bookmarks

- Abstract
 - Background
 - Results
 - Conclusions
 - Background
 - Implementation
 - Results and Discussion
 - Conclusions
- Availability and requirements
 - Prerequisites
 - Acknowledgements
 - Authors' contributions
 - References

Hughes *BMC Bioinformatics* 2011, 12:178
<http://www.biomedcentral.com/1471-2105/12/178>

SOFTWARE

TreeRipper web application: towards a fully automated optical tree recognition

Joseph Hughes

Hughes - 2011 - TreeRipper web application towards a fully automated optical tree recognition

Description	Advanced
Illustrator	
Advanced	<ul style="list-style-type: none"> ▶ PDF Properties (pdf, http://ns.adobe.com/pdf/1.3/) ▶ XMP Core Properties (xmp, http://www.adobe.com/xmp/) ▼ Dublin Core Properties (dc, http://purl.org/dc/terms/) <ul style="list-style-type: none"> dc:format: application/pdf dc:identifier: http://dx.doi.org/10.1186/1471-2105-12-178 ▼ dc:creator (seq container) <ul style="list-style-type: none"> [1]: Joseph Hughes ▼ dc:title (alt container) <ul style="list-style-type: none"> [x-default]: TreeRipper web application: towards a fully automated optical tree recognition ▼ dc:description (alt container) <ul style="list-style-type: none"> [x-default]: BMC Bioinformatics ▼ dc:publisher (bag container) <ul style="list-style-type: none"> [1]: BioMed Central Ltd ▼ dc:date (seq container) <ul style="list-style-type: none"> [1]: 2011-05-20 ▼ dc:rights (alt container) <ul style="list-style-type: none"> [x-default]: © 2011 Hughes ▶ XMP Media Management Properties ▶ XMP Rights Management Properties ▶ http://prismstandard.org/namespaces/1.2/

Powered By **xmp**

Replace... Append...

IMG [Icons]

More Info

Previous Next Zoom

GPS General JFIF Exif TIFF

Latitude 55° 50' 7.2" N
 Longitude 4° 15' 51.6" W



Locate



It's not easy being green!

