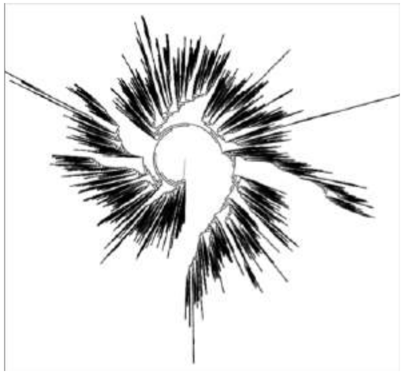
Building, Using, and Tending

The Tree of Life

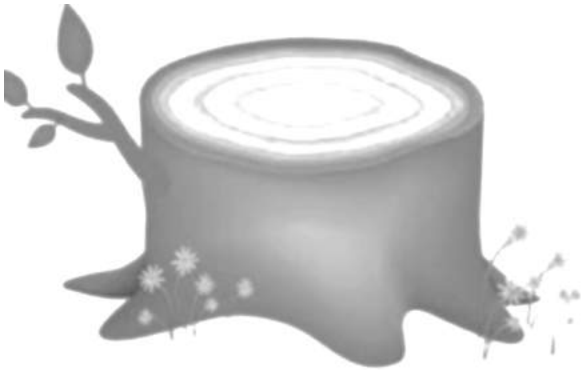
孙苗
Miao Sun

08/22/2018

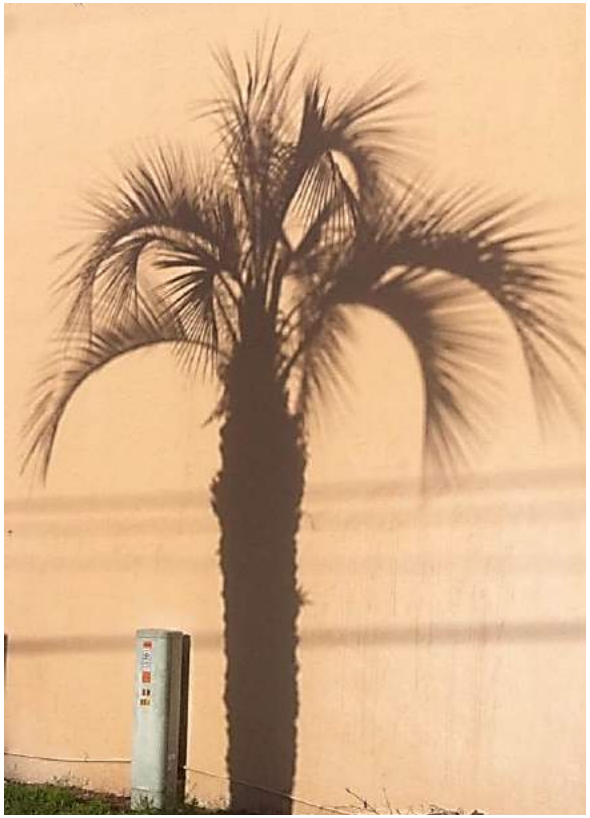


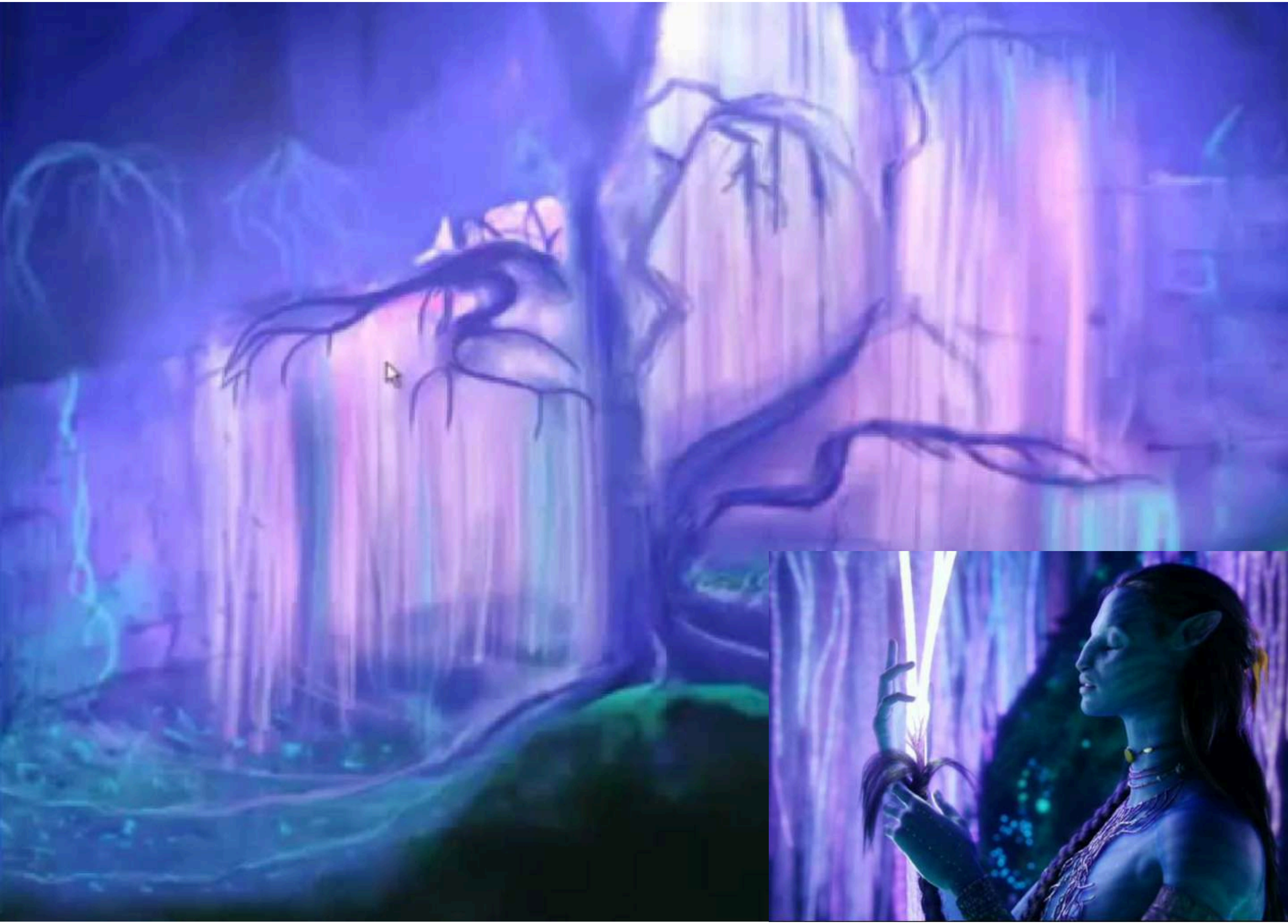
北冥有木，生命之树。树之大，枝枝桠桠数百万年也。

——桩子



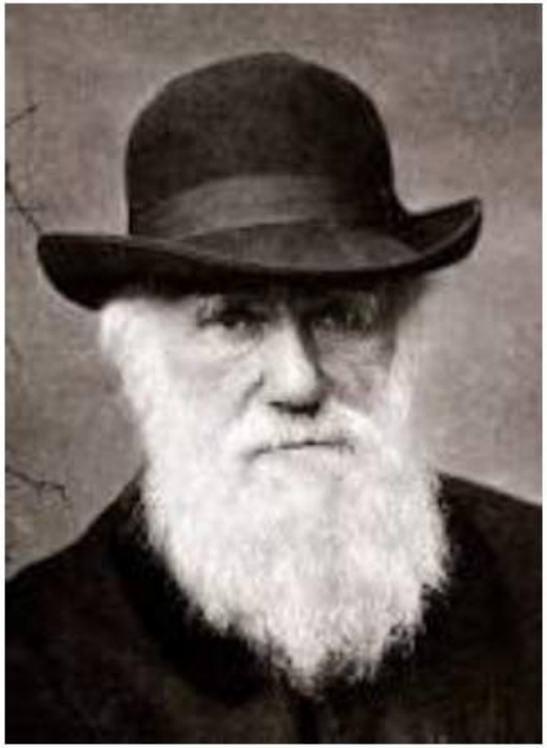
Tree Thinking

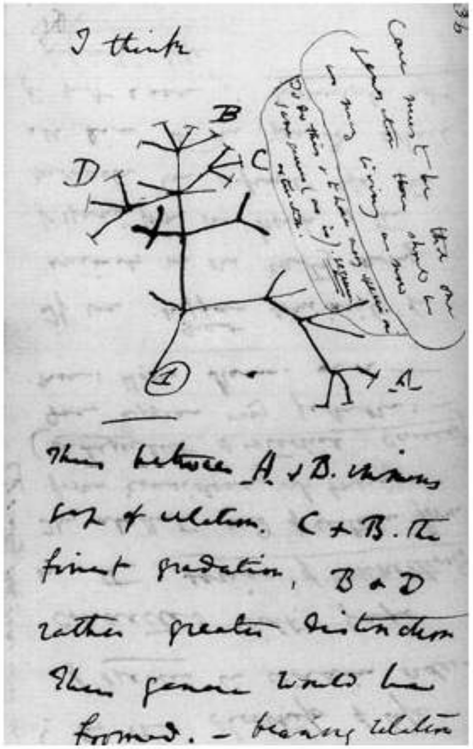




The Concept of Tree of Life

"…and so by generation I believe it has been with the great Tree of Life, which fills with its dead and broken branches the crust of the earth, and covers the surface with its ever-branching and beautiful ramifications" Darwin (1859)

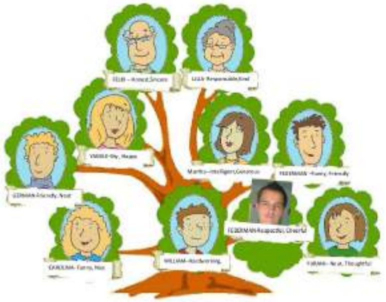


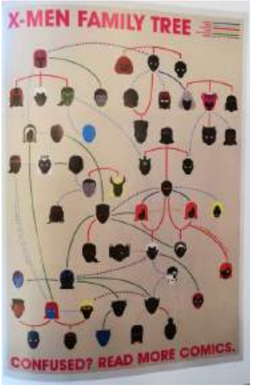


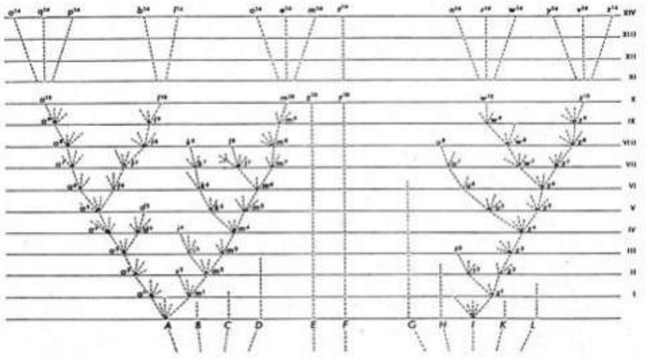
Thus between A & B immens gap of relation. C & B the finest gradation, B & D rather greater distinction Thus genera would be formed. – bearing relation

What is a phylogenetic tree?

- Evolutionary relationships ---like family tree



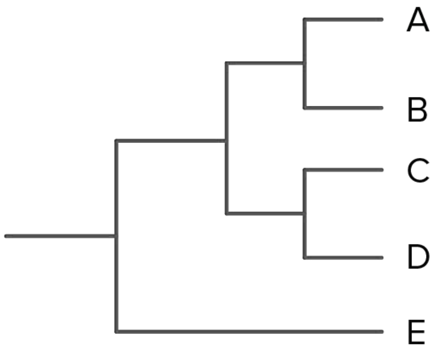


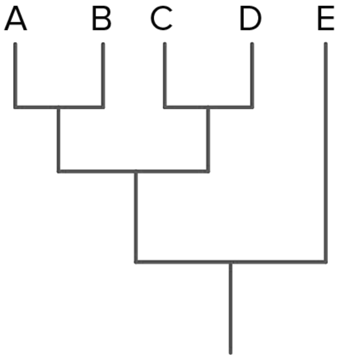


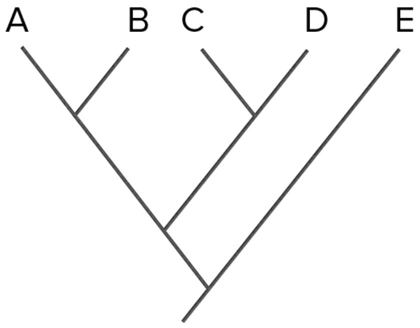
From Darwin's *The Origin of Species* (1859)

What is a phylogenetic tree?

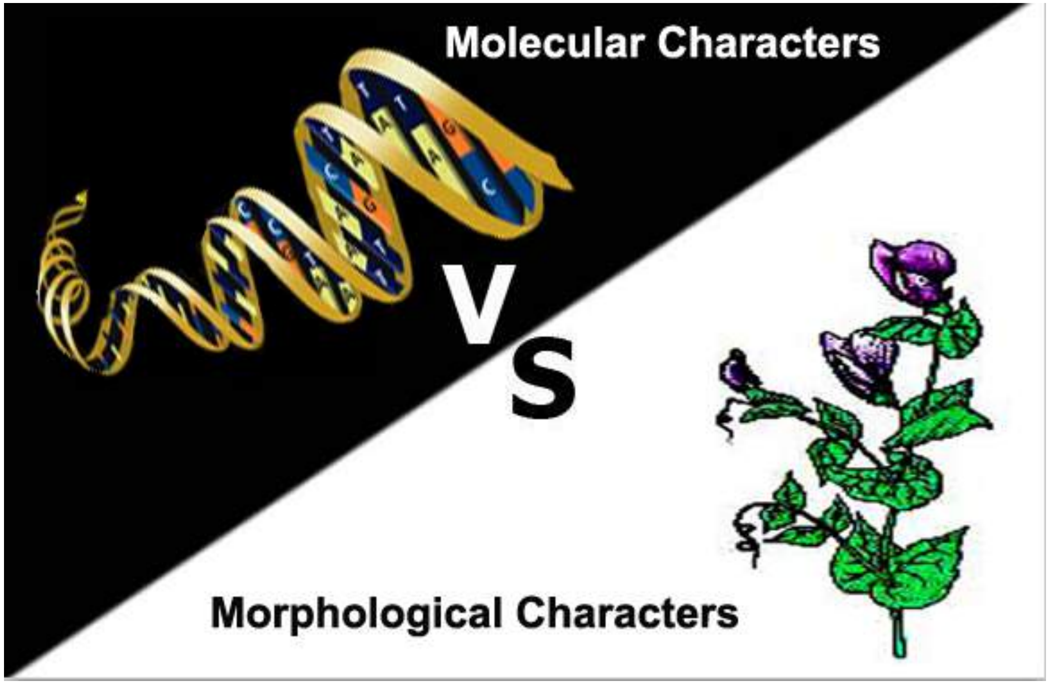
- Branching pattern; common ancestors
- Extinct taxa not include (fossil)
- Various equivalent styles
- Bifurcating tree; complex network
- Hypotheses, not definitive facts







Building the Tree ---data



Homology is easily determined using molecular characters than morphological characters.

Building the Tree ---molecular data

- **Using molecular data (morphological data)**
  - ❖ **Generate data by yourself (loci, genome, etc)**
  - ❖ **<u>Mining public database (e.g., GenBank)</u>**
  - ❖ **Combine**
- **Sampling: Gene Choice and Taxon sampling**
  - ❖ **chloroplast, mitochondrial, and nuclear (plants)**
  - ❖ **representative (different scales)**

生物多样性　2014, **22** (1): 21–39
Biodiversity Science

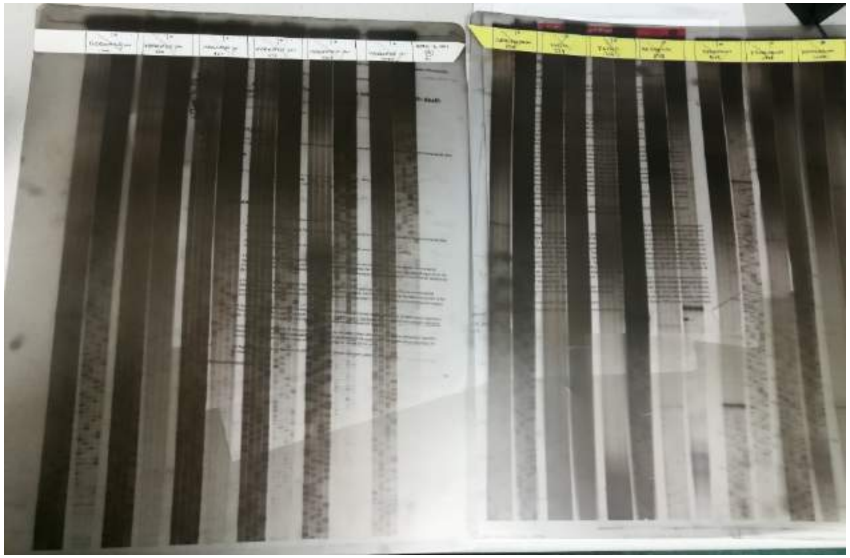
Doi: 10.3724/SP.J.1003.2014.13189
http: //www.biodiversity-science.net

被子植物系统发育深层关系研究：进展与挑战

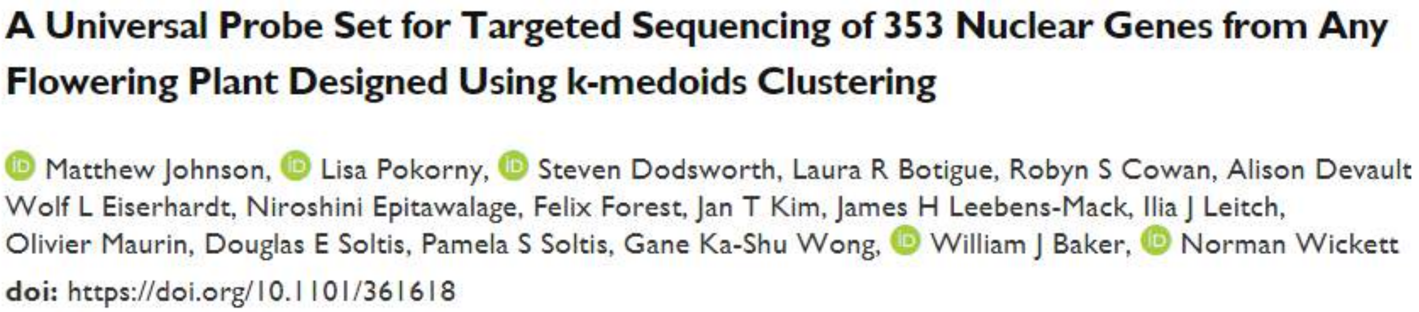
曾丽萍　张　宁　马　红*

(复旦大学生命科学院，遗传工程国家重点实验室，植物科学研究所，进化生物学中心，上海 200433)

Building the Tree



N Generation Sequencing
- **Target Enrichment**



A Universal Probe Set for Targeted Sequencing of 353 Nuclear Genes from Any Flowering Plant Designed Using k-medoids Clustering

Matthew Johnson, Lisa Pokorny, Steven Dodsworth, Laura R Botigue, Robyn S Cowan, Alison Devault Wolf L Eiserhardt, Niroshini Epitawalage, Felix Forest, Jan T Kim, James H Leebens-Mack, Ilia J Leitch, Olivier Maurin, Douglas E Soltis, Pamela S Soltis, Gane Ka-Shu Wong, William J Baker, Norman Wickett

doi: https://doi.org/10.1101/361618

Growth of molecular data for rosids in GenBank (2008-current)

The growth of the nucleotide data for rosids

The growth of the genomic data for rosids

Building the Tree ---Standard procedure

- Alignment (Molecular and Morphological)
- Evolutionary model (substitution rate)
- Inference methods:
  - Maximum Likelihood (ML tree): RAxML
  - Maximum Parsimony (MP tree): PAUP, GARLIC
  - Bayesian Inference (BI tree): MrBayes (Metropolis-coupled Markov Chain [MCMC])

Building the Tree --- software

- Software:
- http://evolution.genetics.washington.edu/phylip/software.html

MEGA Molecular Evolutionary Genetics Analysis

Building the Tree ---online webserver

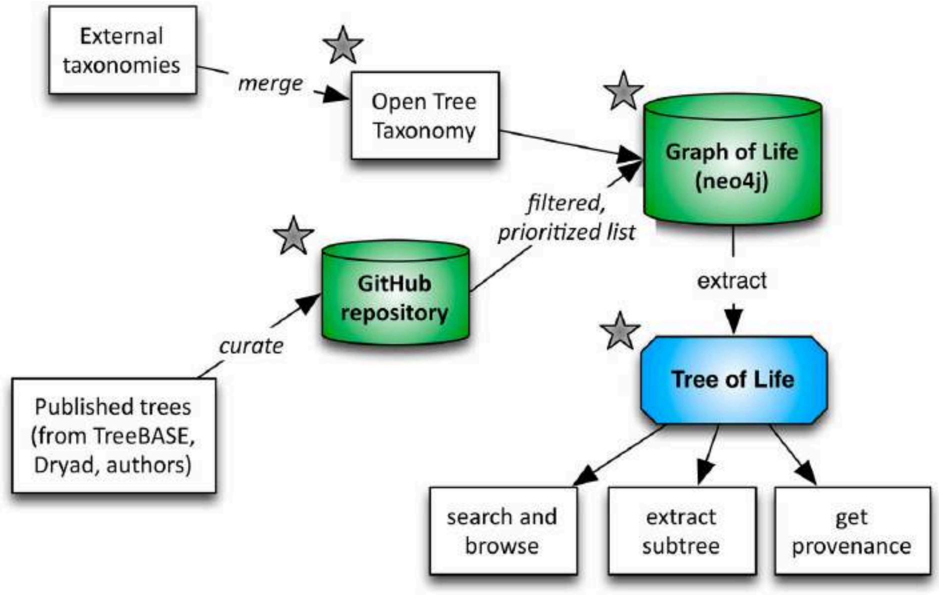
- Web-server:

Phylogeny.fr
Robust Phylogenetic Analysis For The Non-Specialist

  - http://www.phylogeny.fr/
  - Simple Phylogeny: https://www.ebi.ac.uk/Tools/phylogeny/simple_phylogeny/
  - CIPRES: https://www.phylo.org

Building the Tree ---subset supertree

- **Using phylogenetic data**
  - ❖**Built by yourself**
  - ❖<u>Open Tree of Life</u>



OPEN Tree of Life

Hinchliff et al., 2015

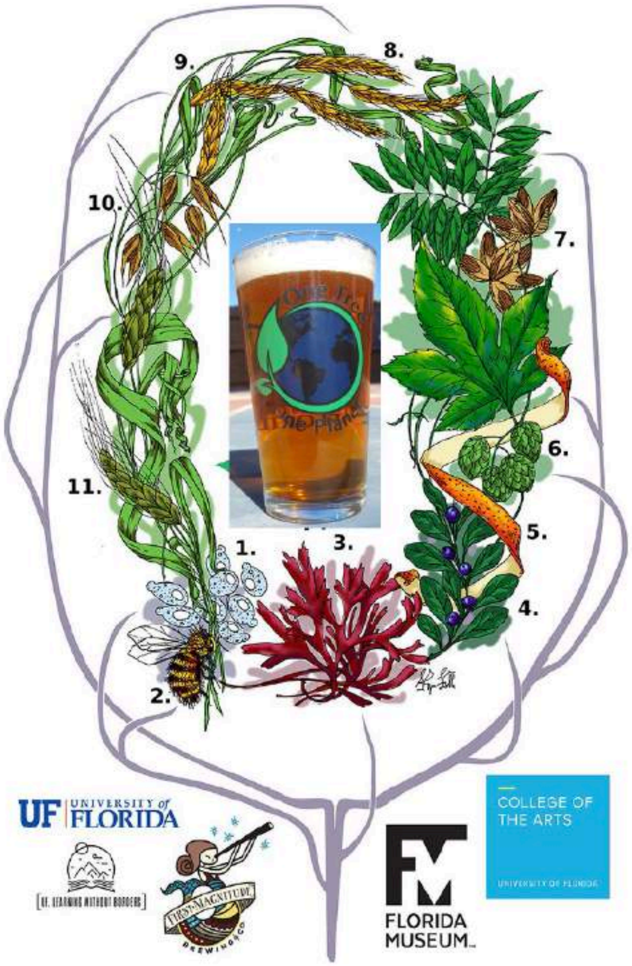
Building the Tree ---Subset tools

- Using published phylogenetic data
- Phylomatic: http://phylodiversity.net/phylomatic/
- http://www.darwintree.cn/flora/index.shtml

- Megatree as backbone
- Target species list
    - fagaceae/quercus/Quercus_robur
    - dipterocarpaceae/Shorea/Shorea_parvifolia
- Format: nexml, newick

- Other software phyx, and newickutils, opentree_pytoy, R packages (e.g., rotl).

Examples 1:




C:/Users/cactus/Desktop/OpentreeExtractor_Shiny-master - Shiny

http://127.0.0.1:3818 | Open in Browser

Open Tree Extractor

Choose csv file to upload

Browse... One_tree_beer.csv

Upload complete

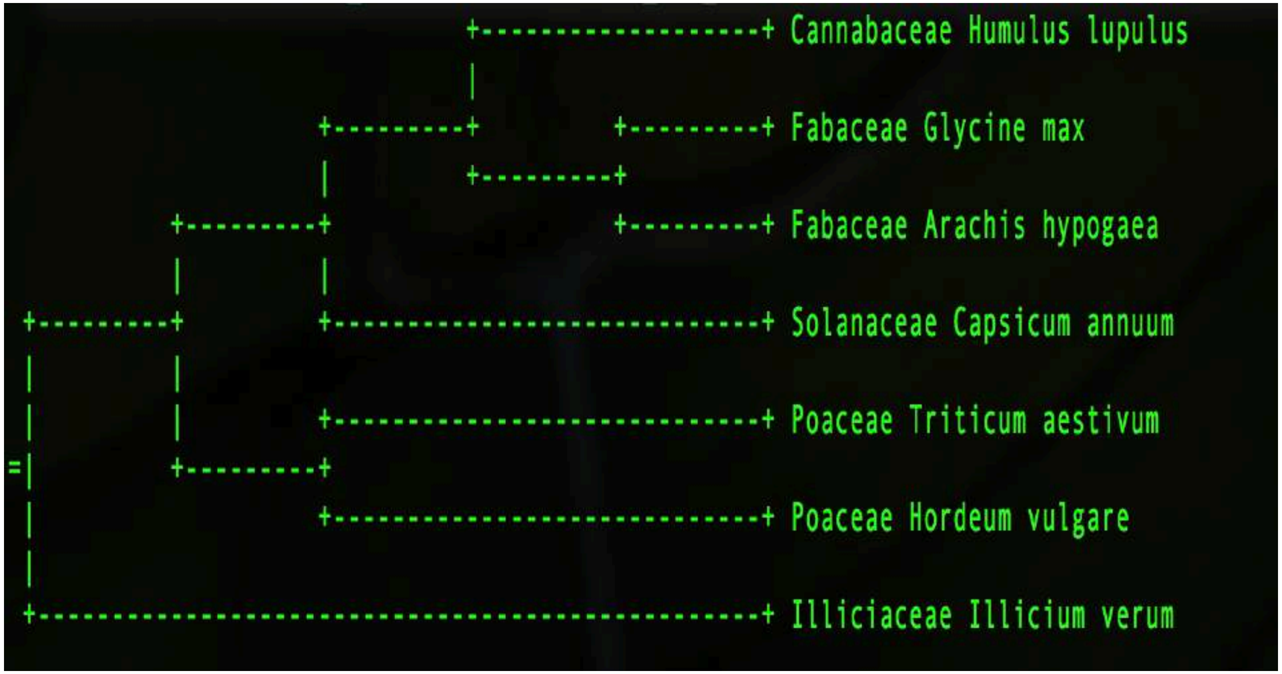
☑ Header

Separator:

● , ○ ; ○ :

Examples 2:

2018
FIFA WORLD CUP
RUSSIA

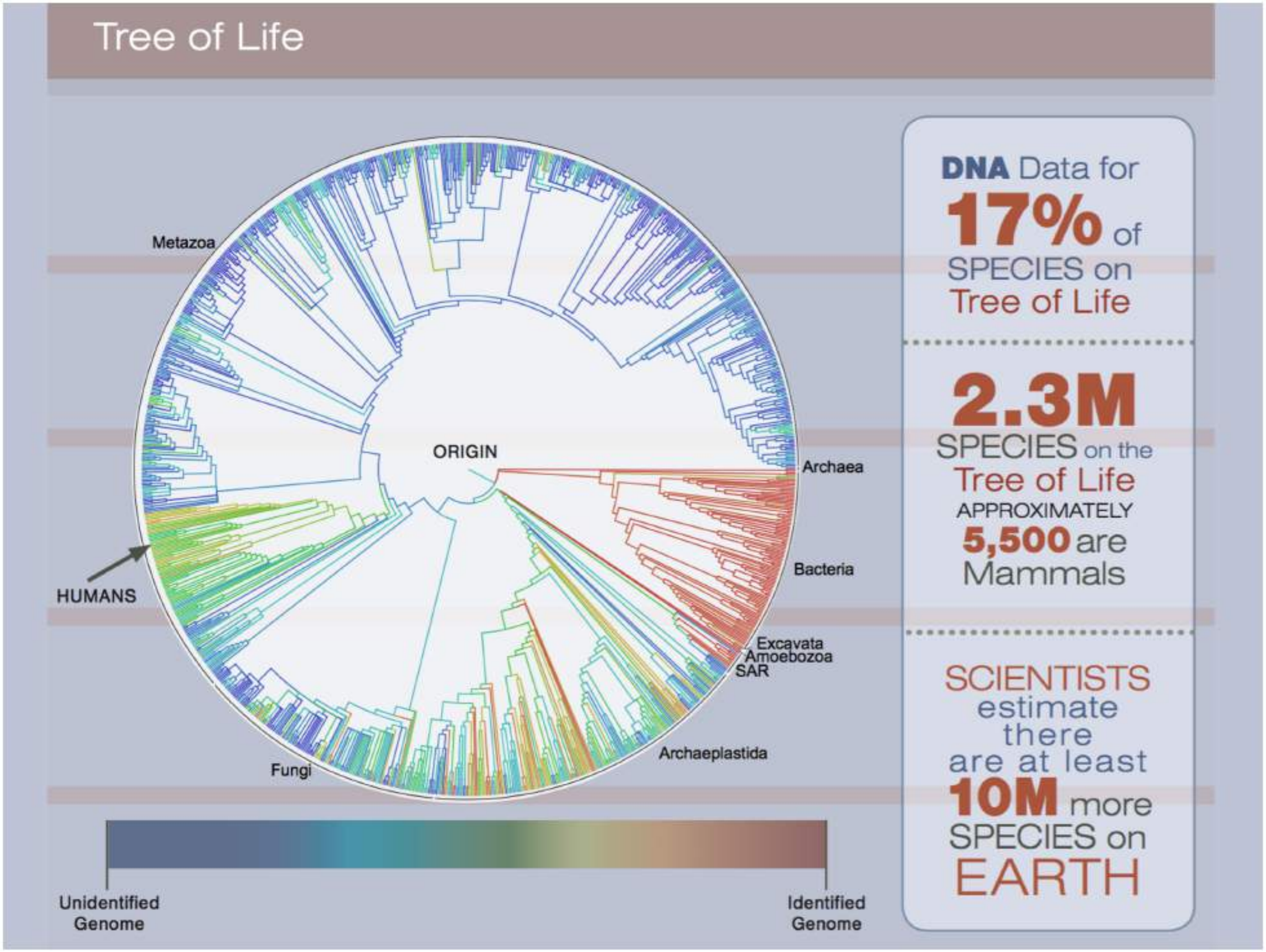


```
                               +-------------------+ Cannabaceae Humulus lupulus
                               |
                     +---------+         +---------+ Fabaceae Glycine max
                     |         +---------+
           +---------+                   +---------+ Fabaceae Arachis hypogaea
           |         |
 +---------+         +-----------------------------+ Solanaceae Capsicum annuum
 |         |
 |         |         +-----------------------------+ Poaceae Triticum aestivum
=|         +---------+
 |                   +-----------------------------+ Poaceae Hordeum vulgare
 |
 +-------------------------------------------------+ Illiciaceae Illicium verum
```

So easy …?!

Still some big Challenges!

Challenge: data availability

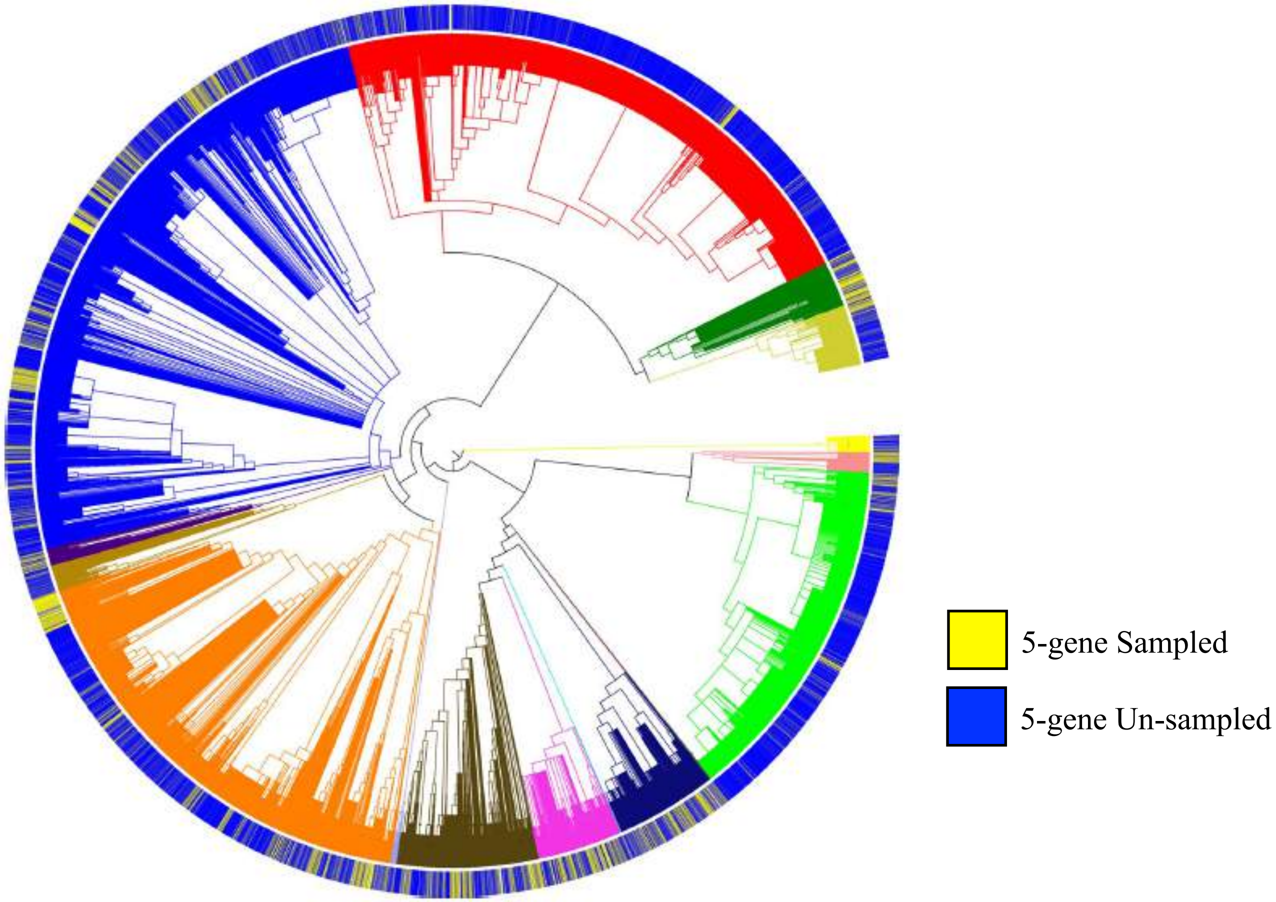


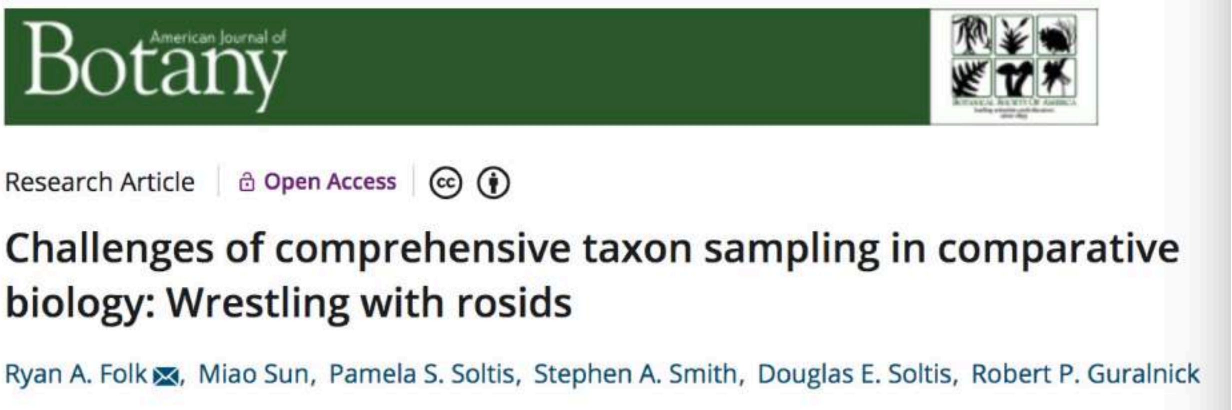
DNA Data for **17%** of SPECIES on Tree of Life

2.3M SPECIES on the Tree of Life APPROXIMATELY **5,500** are Mammals

SCIENTISTS estimate there are at least **10M** more SPECIES on EARTH

data bias



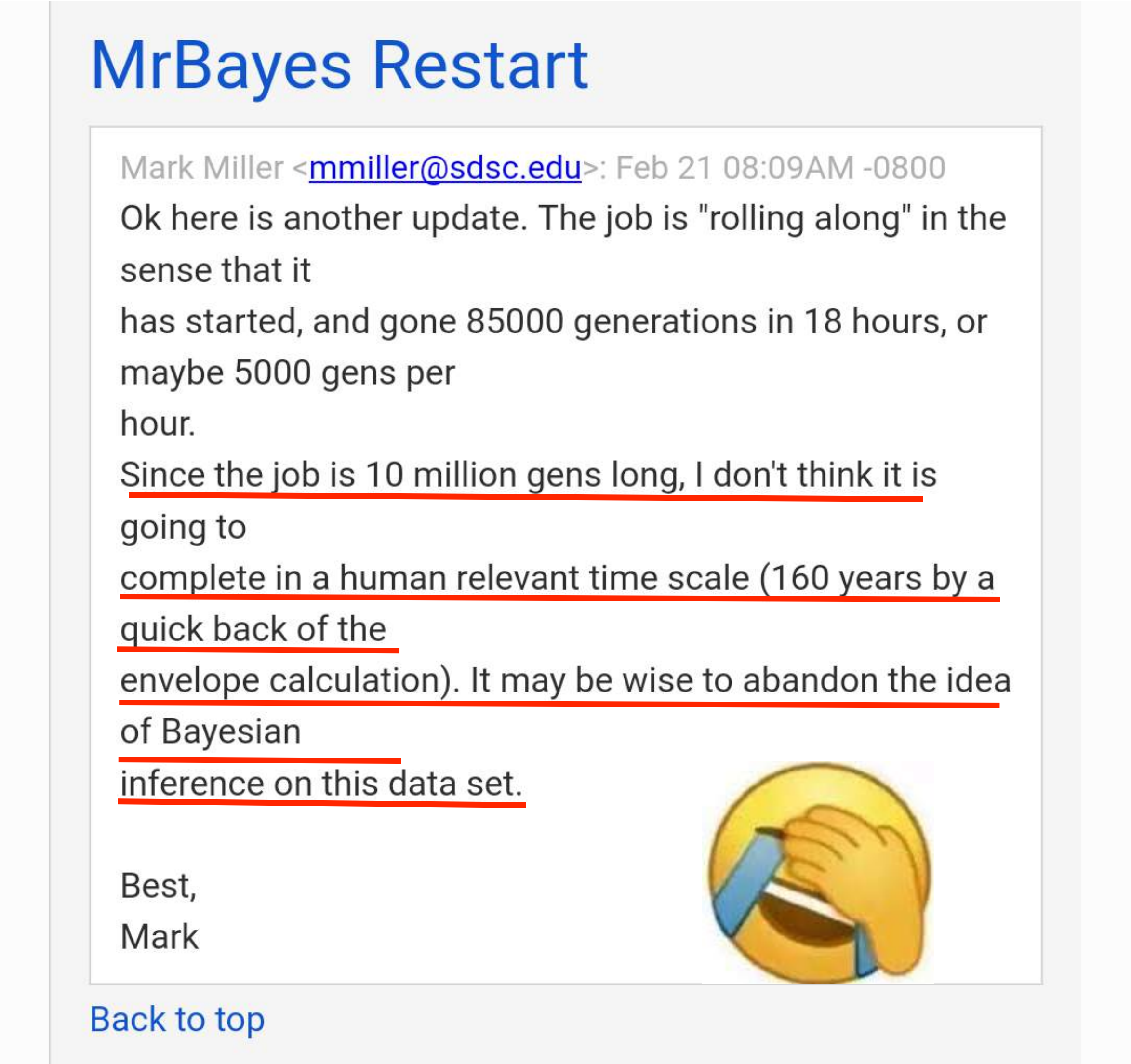


Research Article | Open Access

Challenges of comprehensive taxon sampling in comparative biology: Wrestling with rosids

Ryan A. Folk, Miao Sun, Pamela S. Soltis, Stephen A. Smith, Douglas E. Soltis, Robert P. Guralnick

Challenge: algorithm & Computation



MrBayes Restart

Mark Miller <mmiller@sdsc.edu>: Feb 21 08:09AM -0800

Ok here is another update. The job is "rolling along" in the sense that it
has started, and gone 85000 generations in 18 hours, or maybe 5000 gens per
hour.
Since the job is 10 million gens long, I don't think it is going to
complete in a human relevant time scale (160 years by a quick back of the
envelope calculation). It may be wise to abandon the idea of Bayesian
inference on this data set.

Best,
Mark

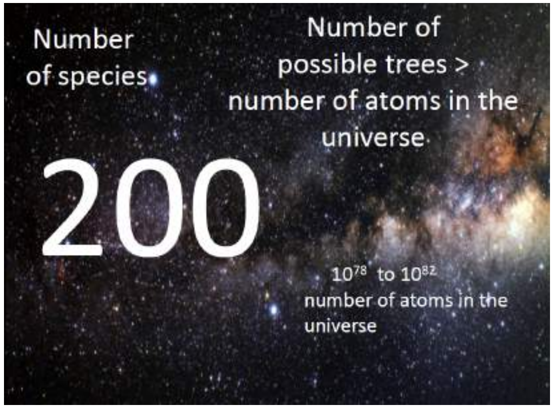
20

SPECIES

6.02x10^{23}

222,000,000,000,000,000,000

POSSIBLE TREES



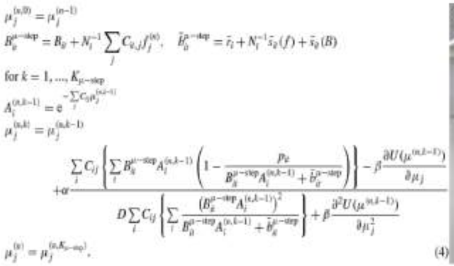
Number of species

200

Number of possible trees > number of atoms in the universe

10^{78} to 10^{82} number of atoms in the universe

- Building big phylogenetic trees a complex problem
- Perfect Storm
  - Algorithm development
  - Computational tools and power
  - DNA sequence revolution



$$\mu_j^{(n,0)} = \mu_j^{(n-1)}$$

$$B_{il}^{\mu\text{-step}} = B_{il} + N_i^{-1}\sum_j C_{ij} f_j^{(n)}, \quad \bar{b}_{il}^{\mu\text{-step}} = \bar{r}_{il} + N_i^{-1}\bar{s}_{il}(f) + \bar{s}_{il}(B)$$

for $k = 1, \ldots, K_{\mu\text{-step}}$

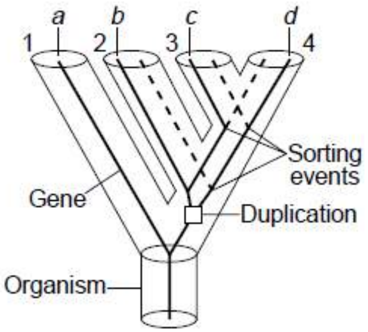
$$A_i^{(n,k-1)} = e^{-\sum_j C_{ij}\mu_j^{(n,k-1)}}$$

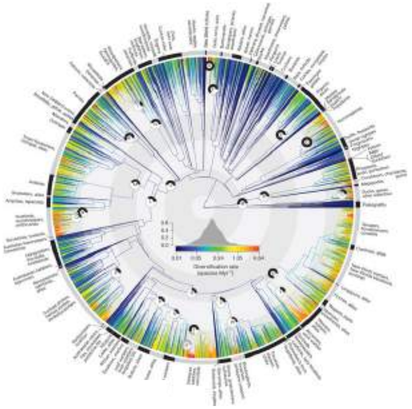
$$\mu_j^{(n,k)} = \mu_j^{(n,k-1)} + \alpha \frac{\sum_i C_{ij}\left\{\sum_l B_{il}^{\mu\text{-step}} A_i^{(n,k-1)}\left(1 - \frac{p_{il}}{B_{il}^{\mu\text{-step}} A_i^{(n,k-1)} + \bar{b}_{il}^{\mu\text{-step}}}\right)\right\} - \beta\frac{\partial U(\mu^{(n,k-1)})}{\partial \mu_j}}{D\sum_i C_{ij}\left\{\sum_l \frac{\left(B_{il}^{\mu\text{-step}} A_i^{(n,k-1)}\right)^2}{B_{il}^{\mu\text{-step}} A_i^{(n,k-1)} + \bar{b}_{il}^{\mu\text{-step}}}\right\} + \beta\frac{\partial^2 U(\mu^{(n,k-1)})}{\partial \mu_j^2}}$$

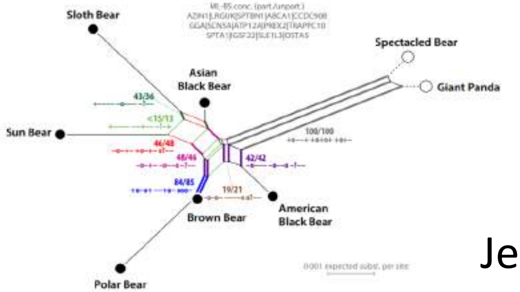
$$\mu_j^{(n)} = \mu_j^{(n,K_{\mu\text{-step}})}. \qquad (4)$$

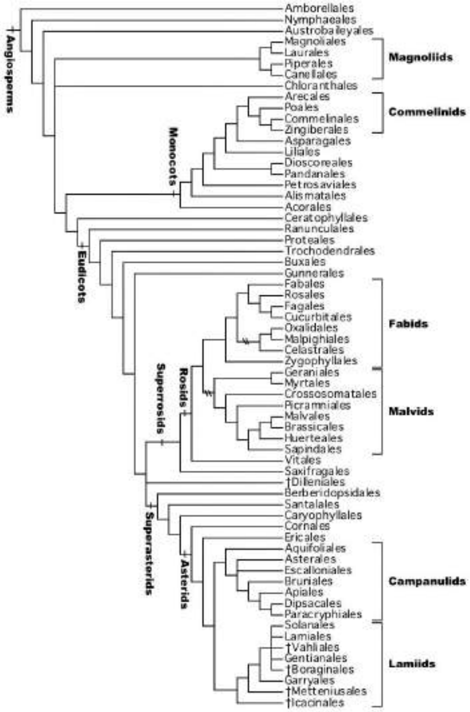
Using the Tree ---research application

- **Phylogenetic framework and taxonomy**
- **Evolutionary events**
- **Diversification analyses**
    - ❖ Diversification Rate shifts
    - ❖ Diversification rate through time plot
    - ❖ Against envormantal/climat profile










生物多样性　2014, **22** (1): 3–20　　Doi: 10.3724/SP.J.1003.2014.13170
Biodiversity Science　　http: //www.biodiversity-science.net

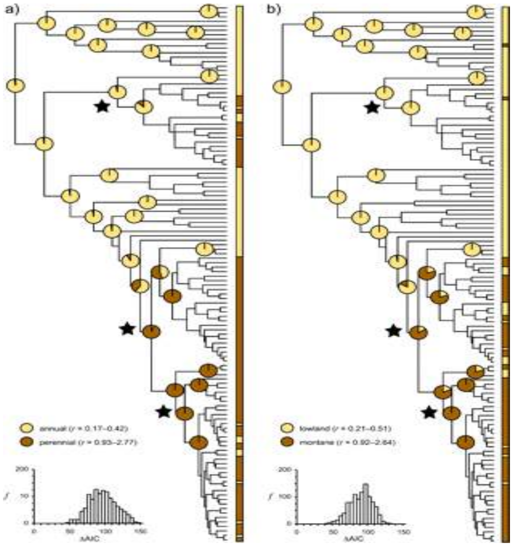
生命之树及其应用

鲁丽敏[1,2]　孙　苗[1,2]　张景博[1,2]　李洪雷[1,2]　林　立[1,2]
杨　拓[1,2]　陈　闽[1,2]　陈之端[1*]

Jetz et al. 2012；　APG IV, 2016; García et al., 2017

Using the Tree ---research application

- **Functional traits**
  - ❖ **Ancestral States Reconstruction**
- **Community ecology, Historical biogeography, Key innovations**
- **<u>Phylogenetic diversity (PD)</u>**
  - ❖ **Distribution data (specimen data)**



###

Comparative tools

- ❖ **R packages (ape, phytools,caper, medusa, geiger, BioGeoBEARs, phangorn, picante, RPANDA, TESS, treepar)**
- ❖ **Python: PyRate, ete3**
- ❖ **Bash: phyx**
- ❖ **C/C++: BAMM**
- ❖ **Rev: RevBayes**

Drummond et al. 2012;

CASE 1:
Tree of Life: China Project

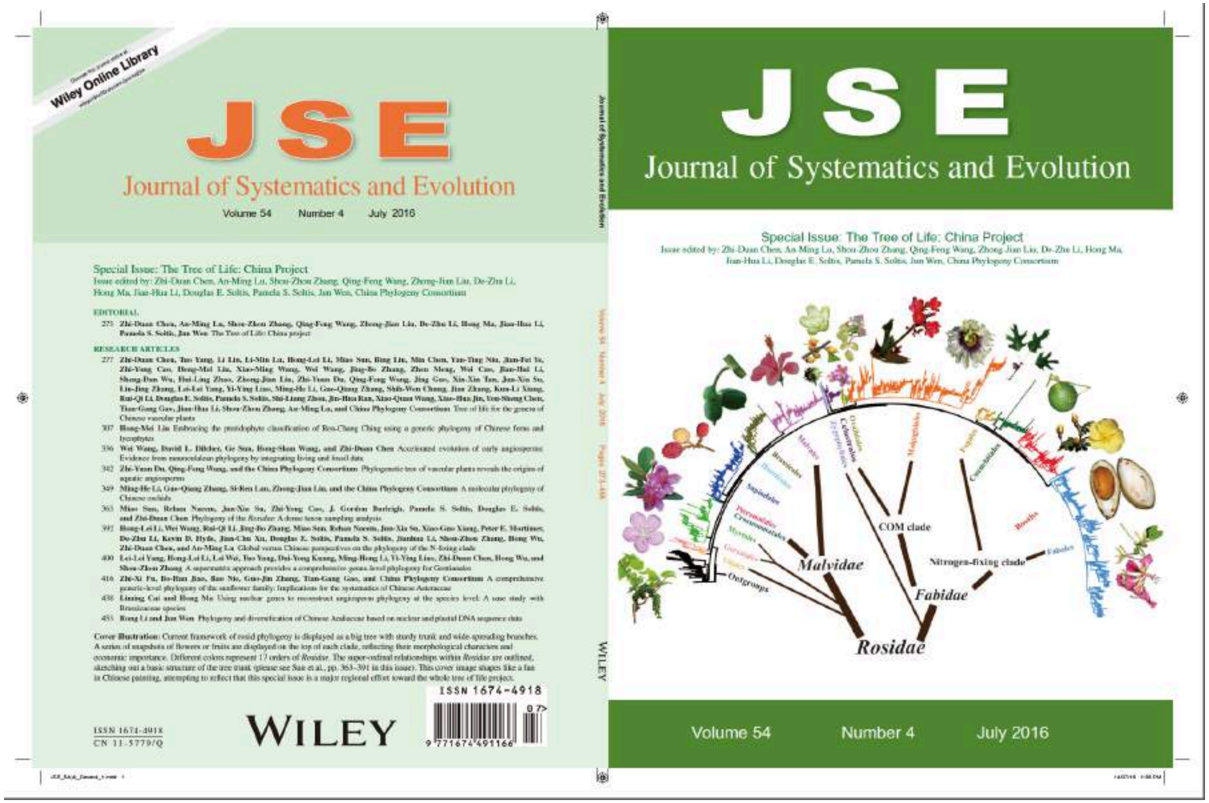
JSE Journal of Systematics and Evolution

doi: 10.1111/jse.12211

Research Article

Phylogeny of the *Rosidae*: A dense taxon sampling analysis

Miao Sun[1,2,3,4], Rehan Naeem[5], Jun-Xia Su[6], Zhi-Yong Cao[1], J. Gordon Burleigh[3,7], Pamela S. Soltis[4,7], Douglas E. Soltis[3,4,7]*, and Zhi-Duan Chen[1]*

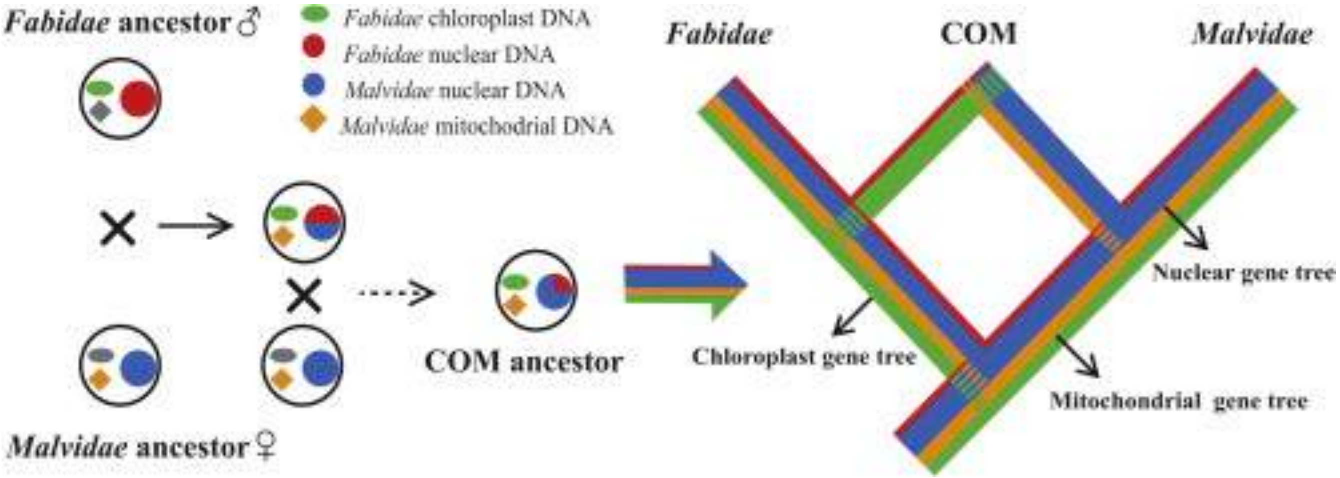


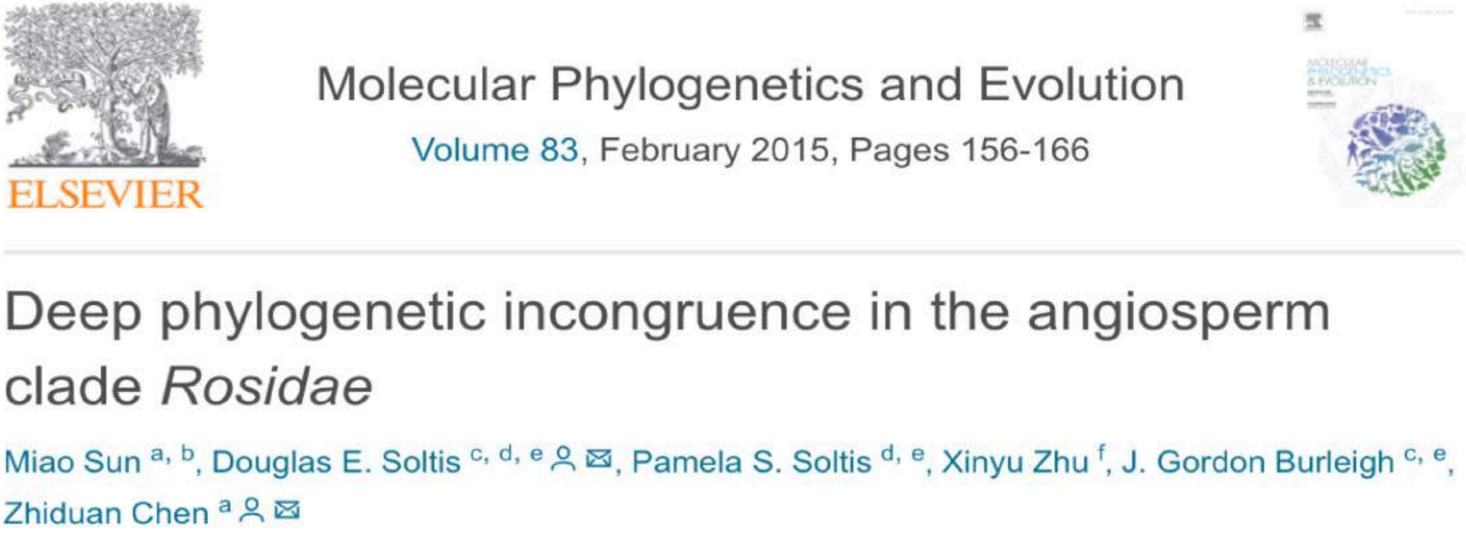
JSE
Journal of Systematics and Evolution

Special Issue: The Tree of Life: China Project
Issue edited by: Zhi-Duan Chen, An-Ming Lu, Shou-Zhou Zhang, Qing-Feng Wang, Zhong-Jian Liu, De-Zhu Li, Hong Ma, Jian-Hua Li, Douglas E. Soltis, Pamela S. Soltis, Jun Wen, China Phylogeny Consortium

Volume 54 Number 4 July 2016

CASE 2:





ELSEVIER

Molecular Phylogenetics and Evolution

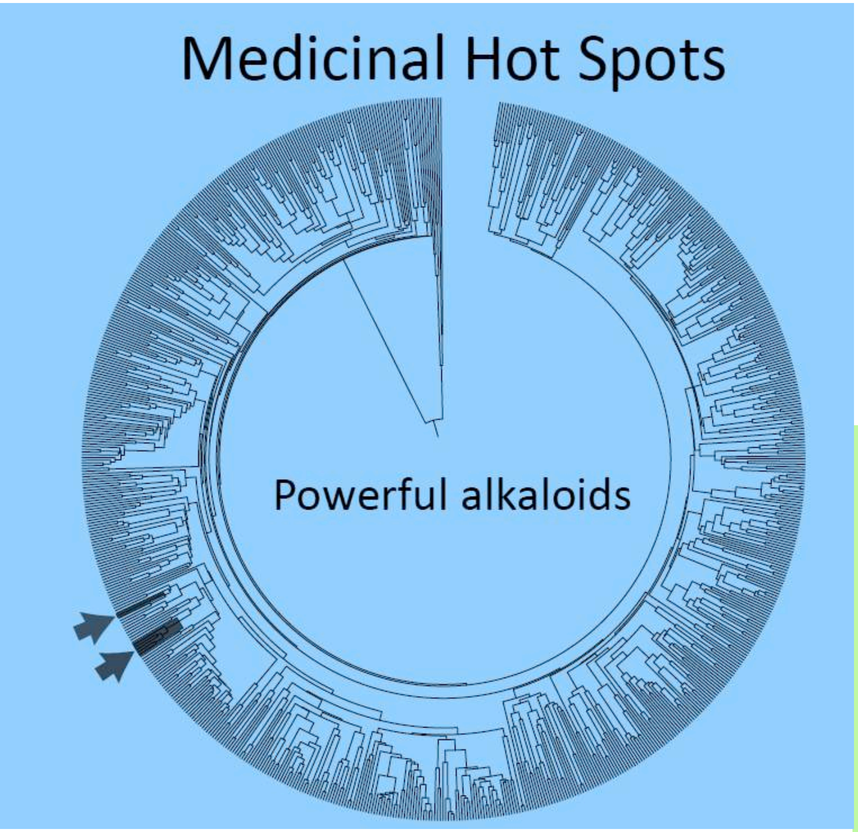
Volume 83, February 2015, Pages 156-166

Deep phylogenetic incongruence in the angiosperm clade *Rosidae*

Miao Sun [a, b], Douglas E. Soltis [c, d, e], Pamela S. Soltis [d, e], Xinyu Zhu [f], J. Gordon Burleigh [c, e], Zhiduan Chen [a]

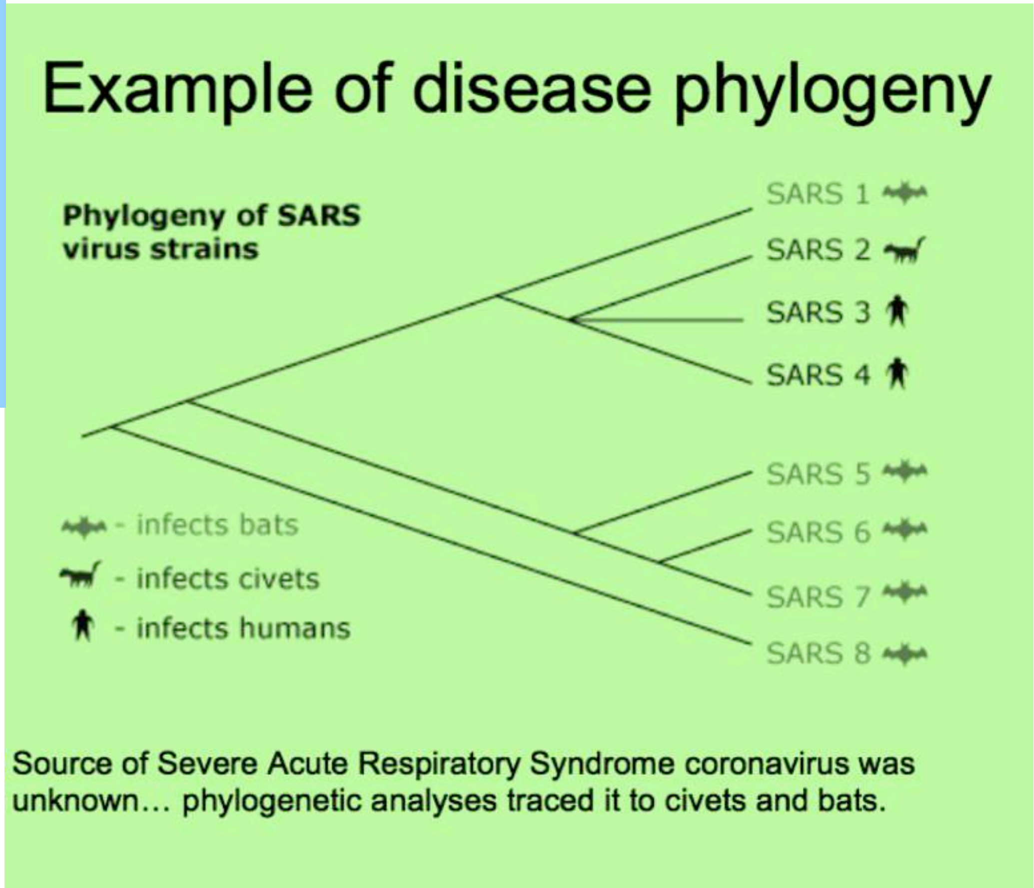
Using the Tree ---more practical

- Biodiversity
- Eco-system
- Conservation
- Climate change
- Crop improvement



Medicinal Hot Spots

- **Disease**
- **Drug Discovery**



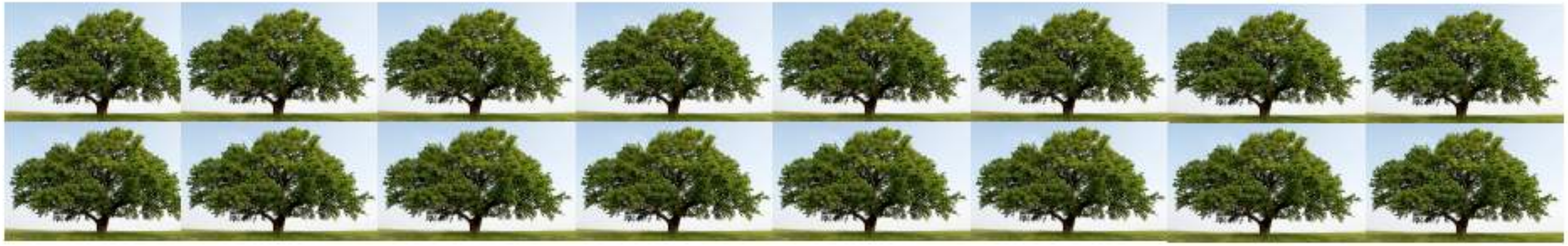
Example of disease phylogeny

Phylogeny of SARS virus strains

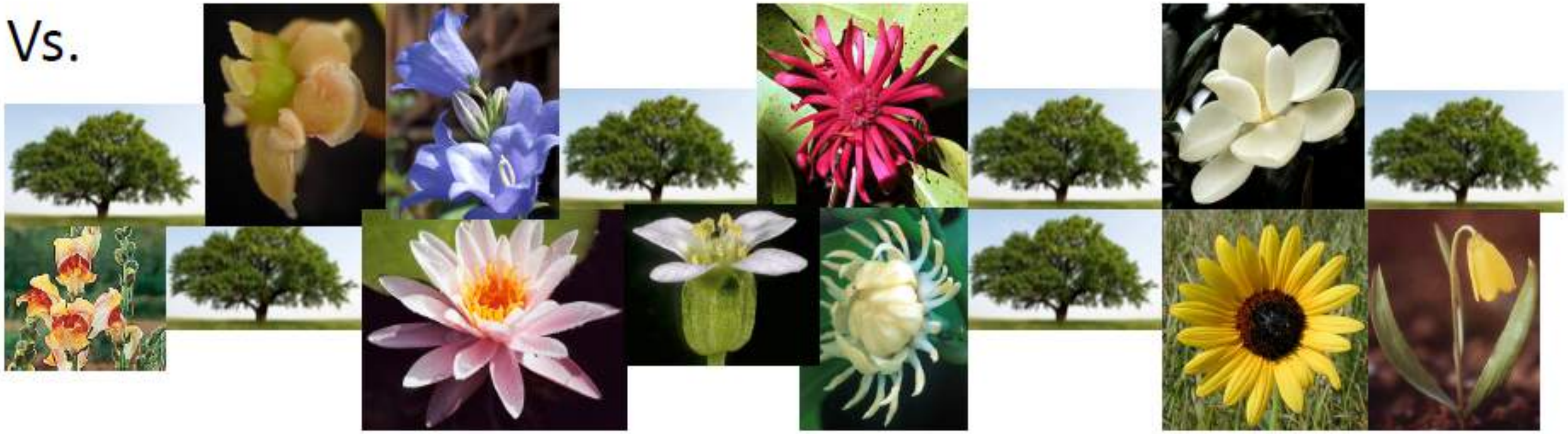
Source of Severe Acute Respiratory Syndrome coronavirus was unknown… phylogenetic analyses traced it to civets and bats.

Phylogenetic diversity (PD)

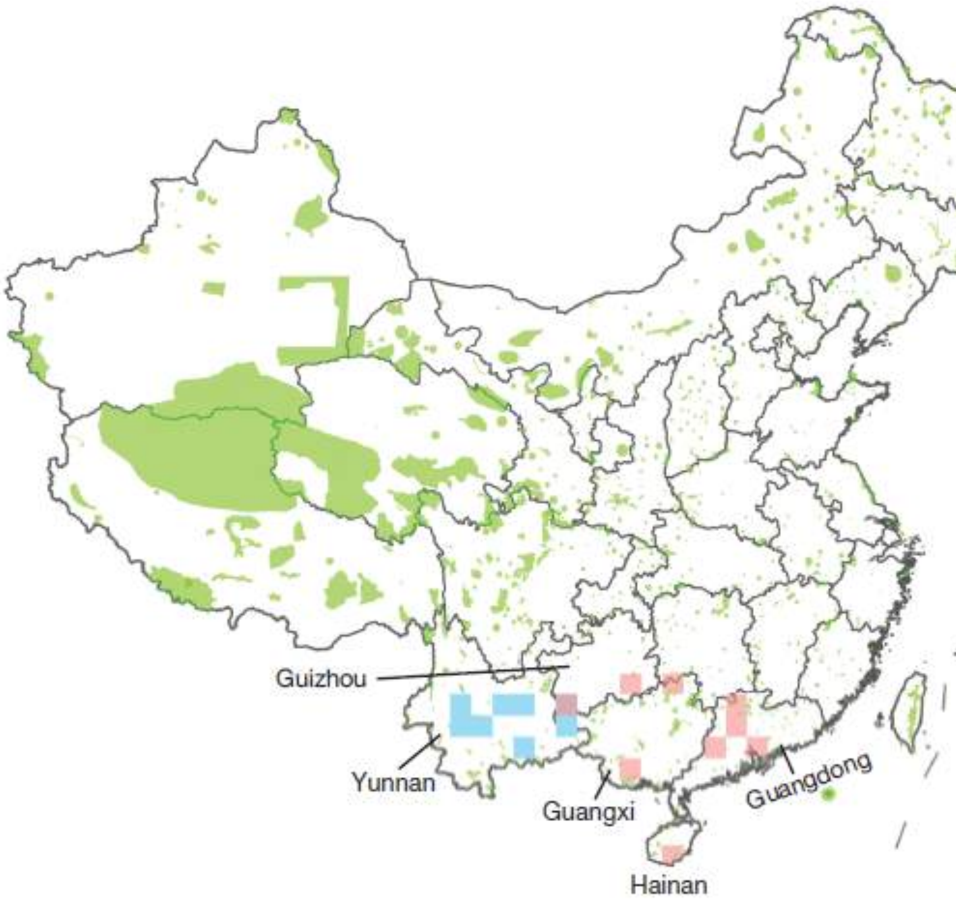
Oaks



Vs.



Nature Reserves Vs. PD hotspots in China



Grid cells with the top 5% highest phylogenetic diversity and SES-PD at genus (pink) and species (blue) levels. Protected areas are highlighted in green. Maps adapted from National Administration of Surveying, Mapping and Geoinformation of China (http://www.sbsm.gov. cn; review drawing number: GS(2016)1576).


nature
International journal of science

Evolutionary history of the angiosperm flora of China

Li-Min Lu, Ling-Feng Mao, Tuo Yang, Jian-Fei Ye, Bing Liu, Hong-Lei Li, Miao Sun, Joseph T. Miller, Sarah Mathews, Hai-Hua Hu, Yan-Ting Niu, Dan-Xiao Peng, You-Hua Chen, Stephen A. Smith, Min Chen, Kun-Li Xiang, Chi-Toan Le, Viet-Cuong Dang, An-Ming Lu, Pamela S. Soltis, Douglas E. Soltis, Jian-Hua Li & Zhi-Duan Chen ✉

Preserving and Teaching the Tree of Life

"Look deep into nature, and then you will understand everything better "

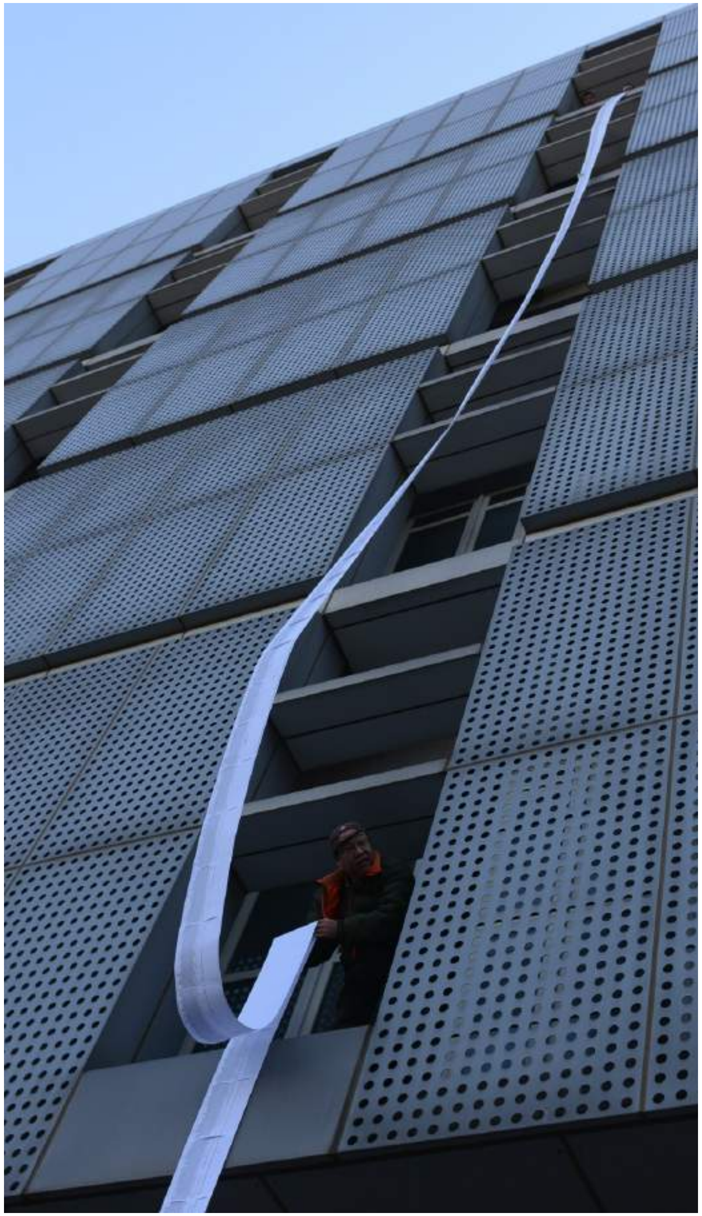
--- Albert Einstein

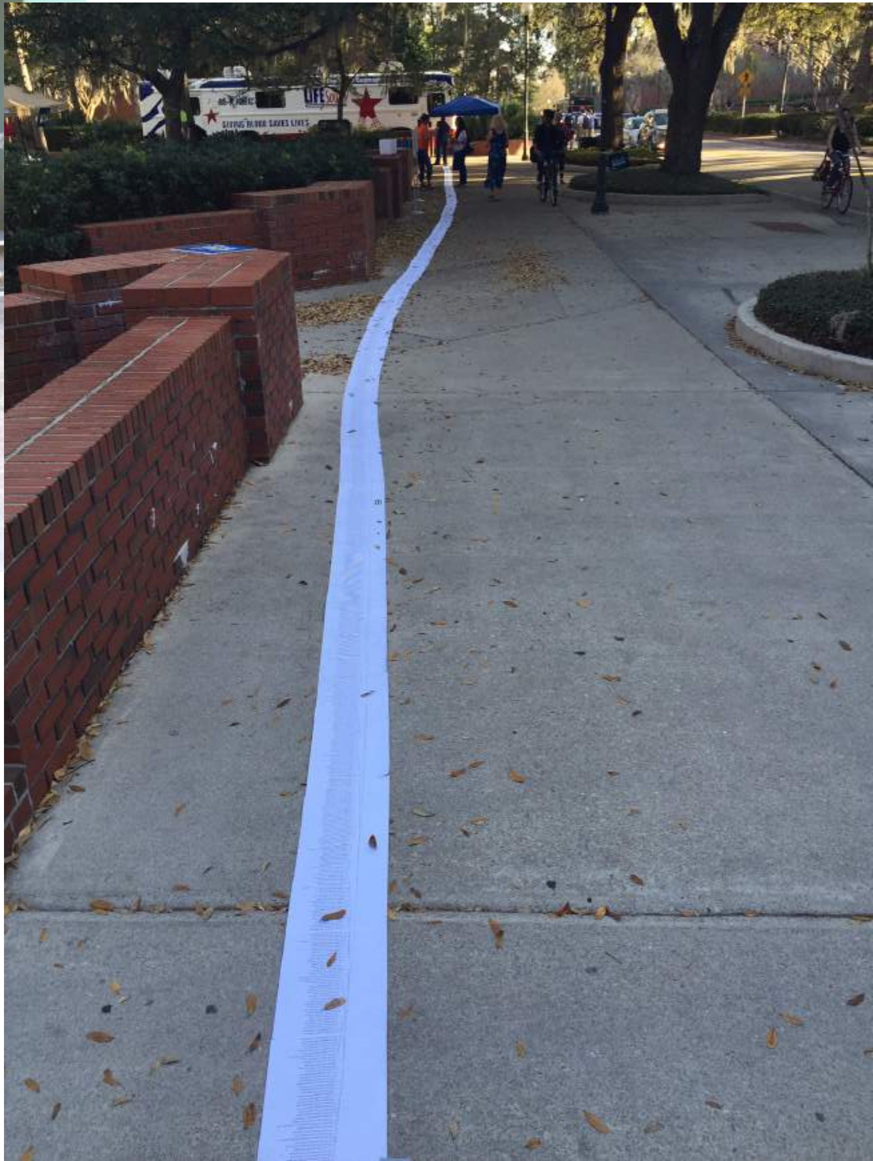
One Tree

One Planet

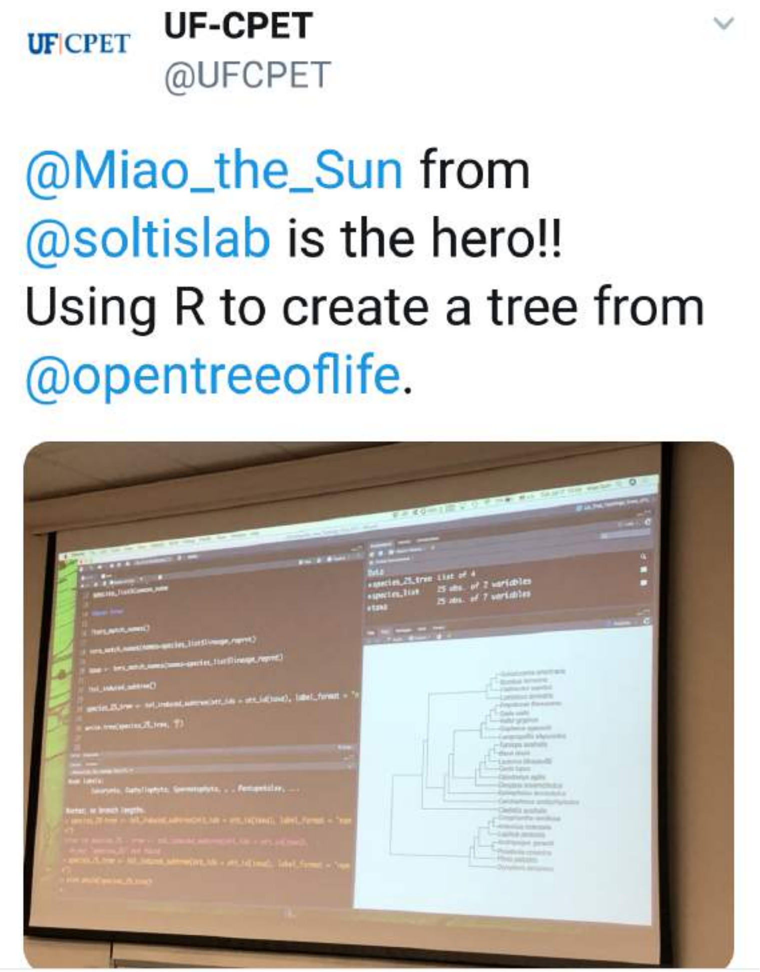

EXIT
EXIT

41 m





LIFE SOUTH
GIVING BLOOD SAVES LIVES



UF-CPET
@UFCPET

@Miao_the_Sun from @soltislab is the hero!! Using R to create a tree from @opentreeoflife.

TreeTender

TREETENDER

THE UNIVERSITY OF FLORIDA

FLORIDA MUSEUM OF NATURAL HISTORY

https://www.treetender.org/

Acknowledgements

Solti Lab

Soltis Lab @UF

@soltislab Follows you

The Soltis Lab @UF @Floridamuseum Laboratory of Molecular Systematics and Evolutionary Genetics. Tweets by @m_gitz and guests

Gainesville, FL

floridamuseum.ufl.edu/museum-voices/...

196 Following **340** Followers

Zhiduan Chen’s Lab