Explore the Angiosperm Diversity Using the Tree of Life and Biological Big Data

利用生命之树与生物学大数据探索被子植物多样性

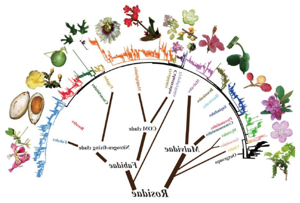
Miao Sun | 孙苗

果蔬园艺作物种质创新与利用全国重点实验室
National Key Laboratory for Germplasm Innovation & Utilization of Horticultural Crops

2025.8.24-8.26 | 哈尔滨

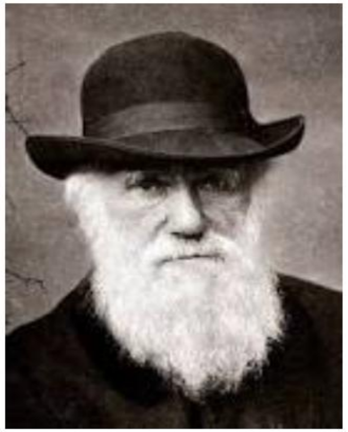
Outlines

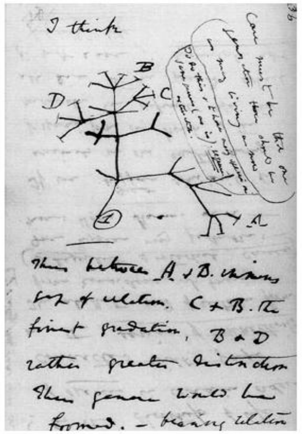
- Tree of Life
- Biological Big Data
- Applications of Tree of Life (My research cases)

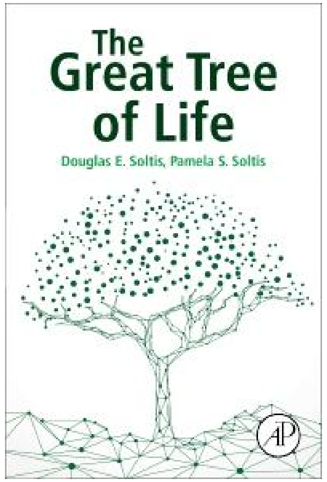


1. Tree of Life

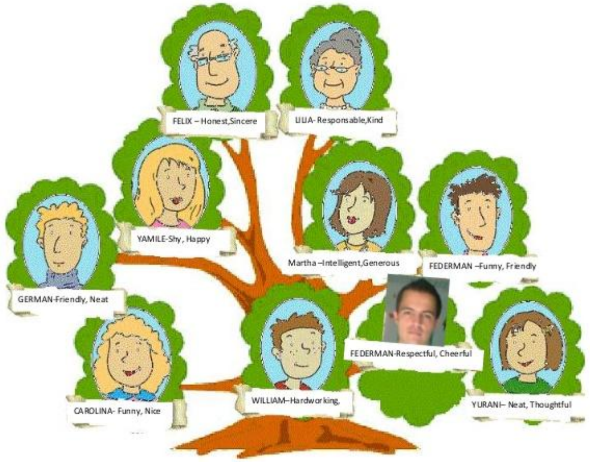
"…and so by generation I believe it has been with the great Tree of Life, which fills with its dead and broken branches the crust of the earth, and covers the surface with its ever-branching and beautiful ramifications" Darwin (1859)







Soltis & Soltis, 2019



Phylogenetic tree

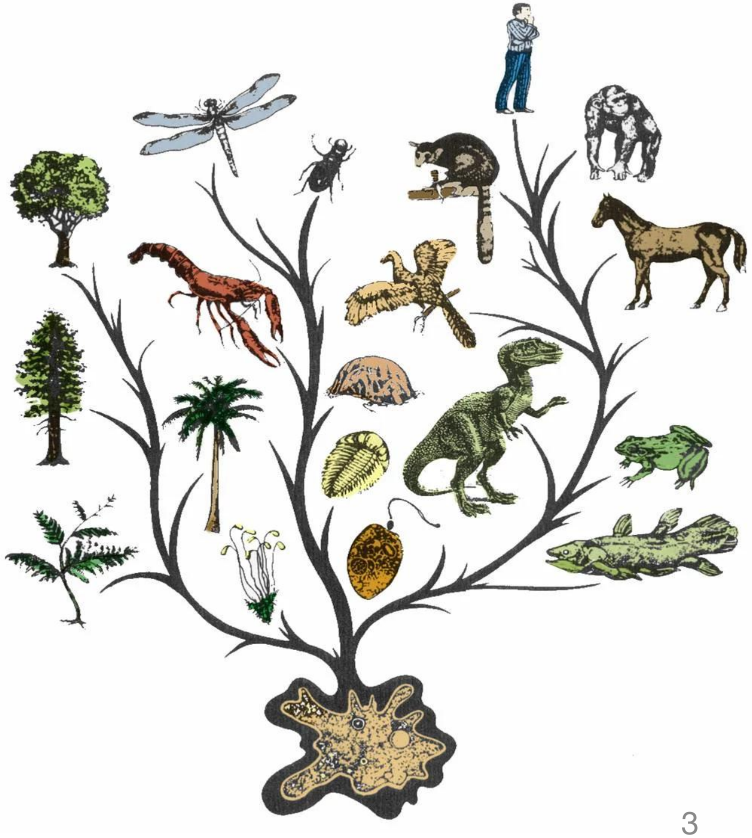
系统发育树

"Family tree"

tree-like topologies to describe the relationships

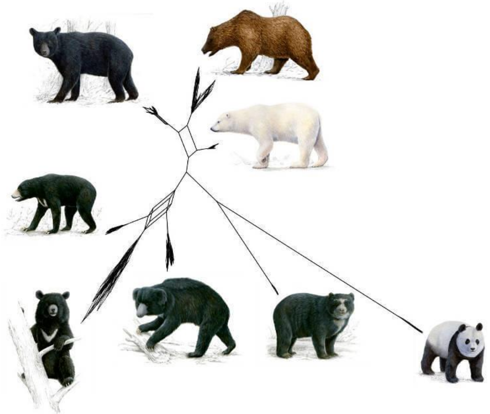
生命之树

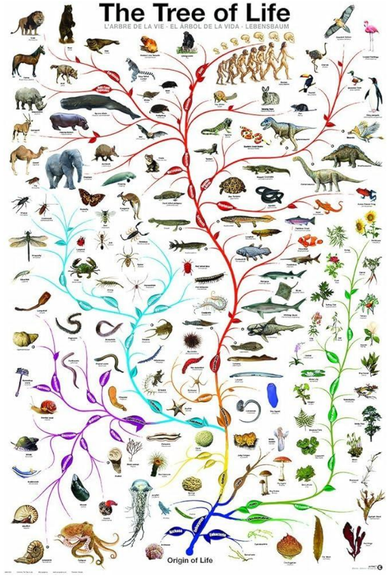
Tree of Life

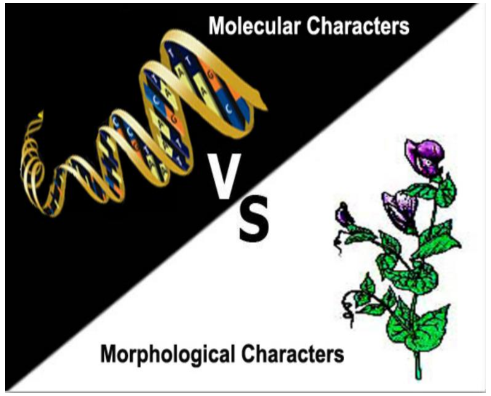


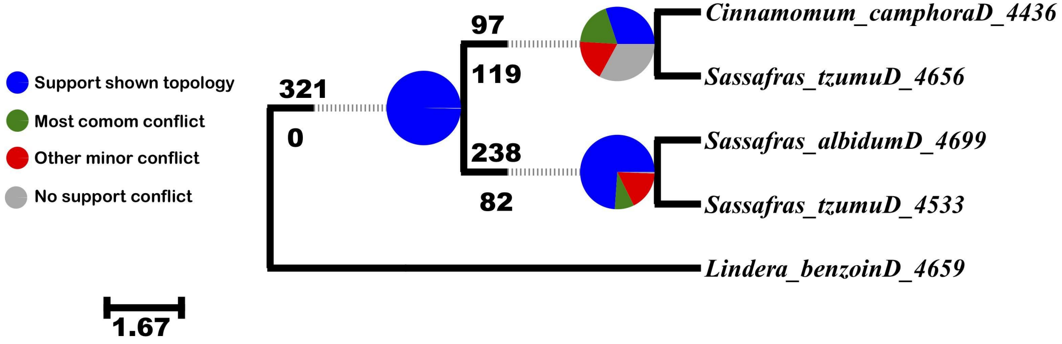
Tree of Life

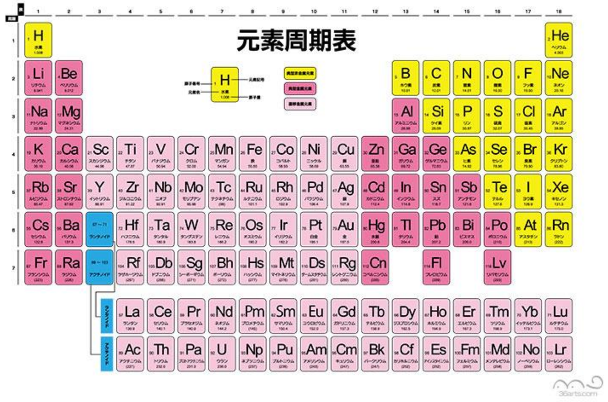
- Root/unroot; branching pattern; common ancestors
- Extinct taxa not include (fossil)
- Various equivalent styles
- Bipartition tree; complex network
- Hypotheses, not definitive facts



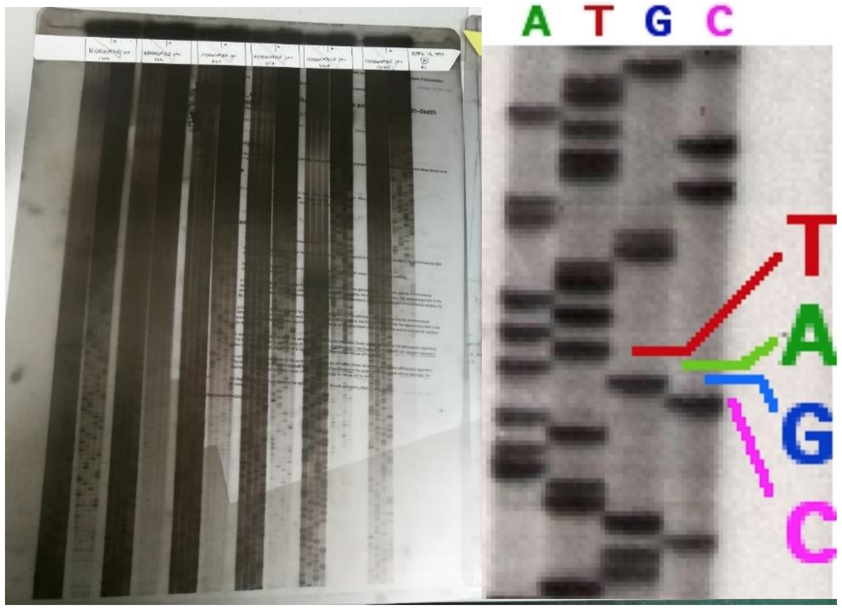


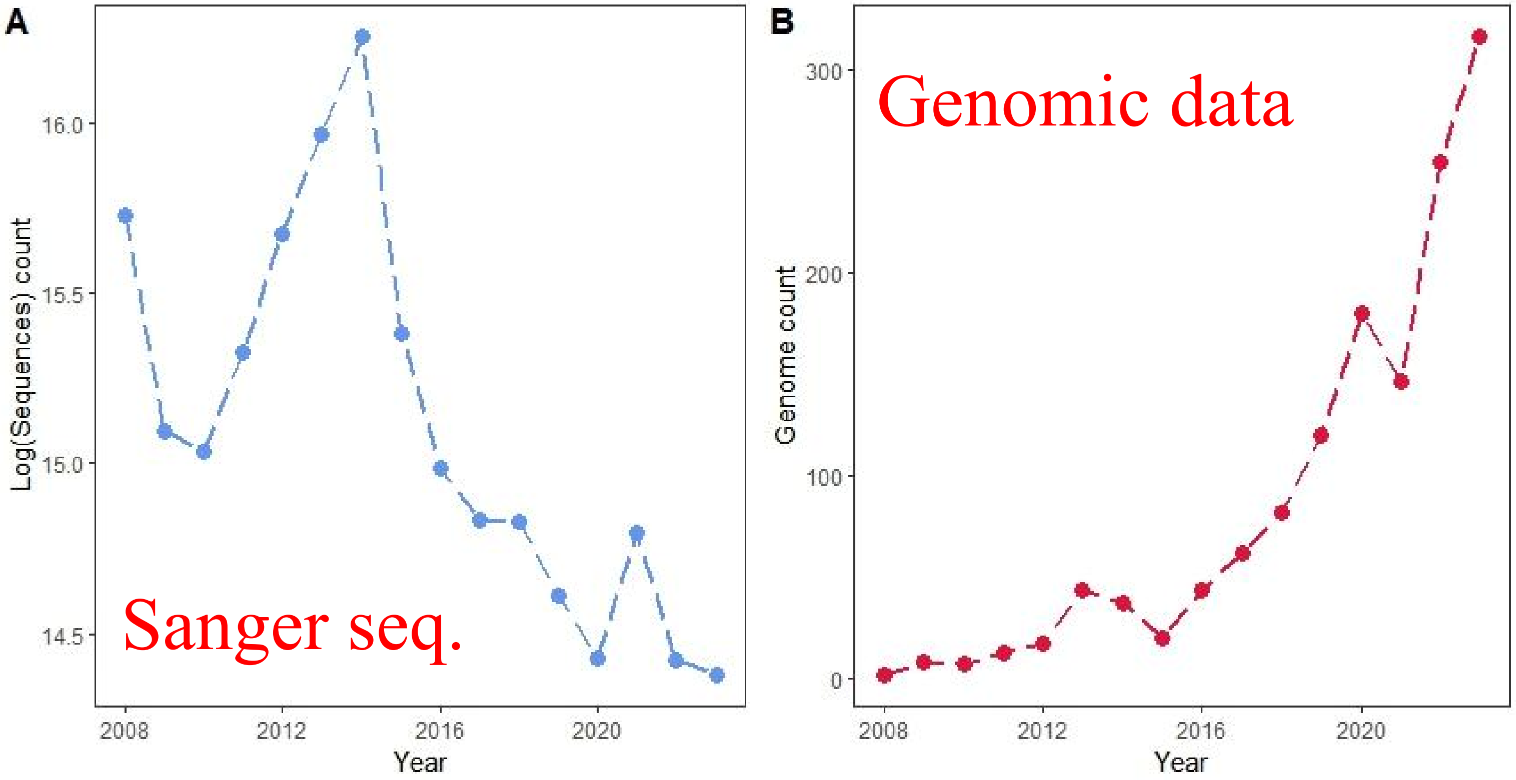






Trend of Sequence Generation





Wang et al.（2024）

A few loci ---->Genome skimming ----> Transcriptomes ----> WGS

Plastid gene/Mitochondrial (maternal) ----> Nuclear genes (biparental)

Dawn of the Nuclear Phylogenomics

华中农业大学 HUAZHONG AGRICULTURAL UNIVERSITY

nature

Explore content About the journal Publish with us

nature > articles > article

Article | Open access | Published: 23 October 2019

One thousand plant transcriptomes and the phylogenomics of green plants

One Thousand Plant Transcriptomes Initiative

Nature 574, 679–685 (2019) | Cite this article

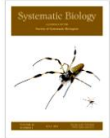
Abstract

Green plants (Viridiplantae) include around 450,000–500,000 species[1,2] of great diversity and have important roles in terrestrial and aquatic ecosystems. Here, as part of the One Thousand Plant Transcriptomes Initiative, we sequenced the vegetative transcriptomes of 1,124 species that span the diversity of plants in a broad sense (Archaeplastida), including green plants (Viridiplantae), glaucophytes (Glaucophyta) and red algae (Rhodophyta). Our analysis provides a robust phylogenomic framework for examining the evolution of green plants. Most inferred species relationships are well supported across multiple species tree and supermatrix analyses, but discordance among plastid and nuclear gene trees at a few important nodes highlights the complexity of plant genome evolution, including polyploidy, periods of rapid speciation, and extinction. Incomplete sorting of ancestral variation, polyploidization and massive expansions of gene families punctuate the evolutionary history of green plants. Notably, we find that large expansions of gene families preceded the origins of green plants, land plants and vascular plants, whereas whole-genome duplications are inferred to have occurred repeatedly throughout the evolution of flowering plants and ferns. The increasing availability of high-quality plant genome sequences and advances in functional genomics are enabling research on genome evolution across the green tree of life.

Leebens-Mack et al.（2019）

Systematic Biology

Issues More content ▾ Submit ▾ Purchase About ▾ Join SSB Systematic Biology

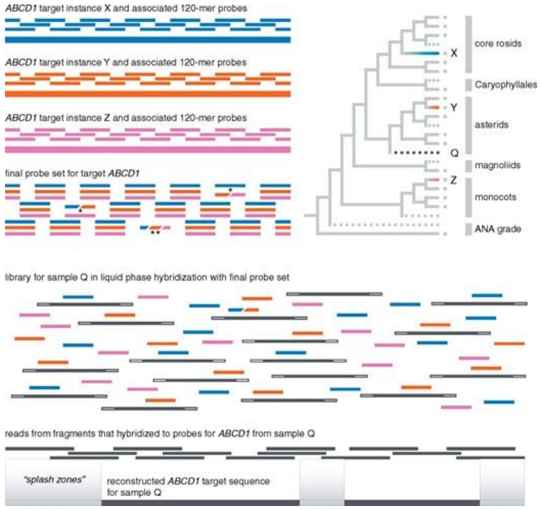


Volume 68, Issue 4
July 2019

JOURNAL ARTICLE

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

Matthew G Johnson, Lisa Pokorny, Steven Dodsworth, Laura R Botigué, Robyn S Cowan, Alison Devault, Wolf L Eiserhardt, Niroshini Epitawalage, Félix 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 J Wickett



Johnson et al.（2019）

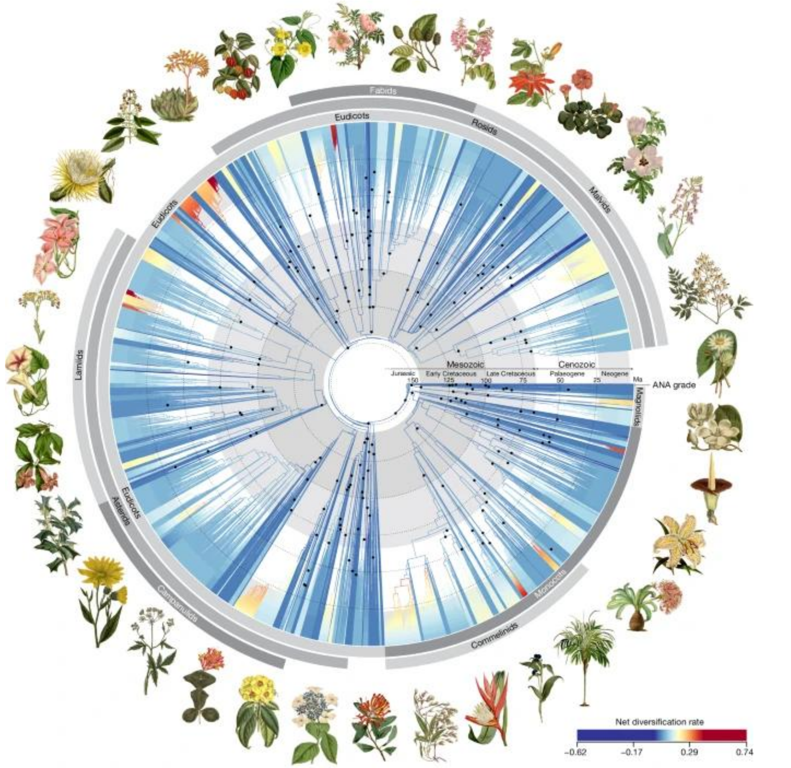
nature

Explore content About the journal Publish with us

nature > articles > article

Article | Open access | Published: 24 April 2024

Phylogenomics and the rise of the angiosperms



Zuntini et al.（2024）

Mega-phylogeny

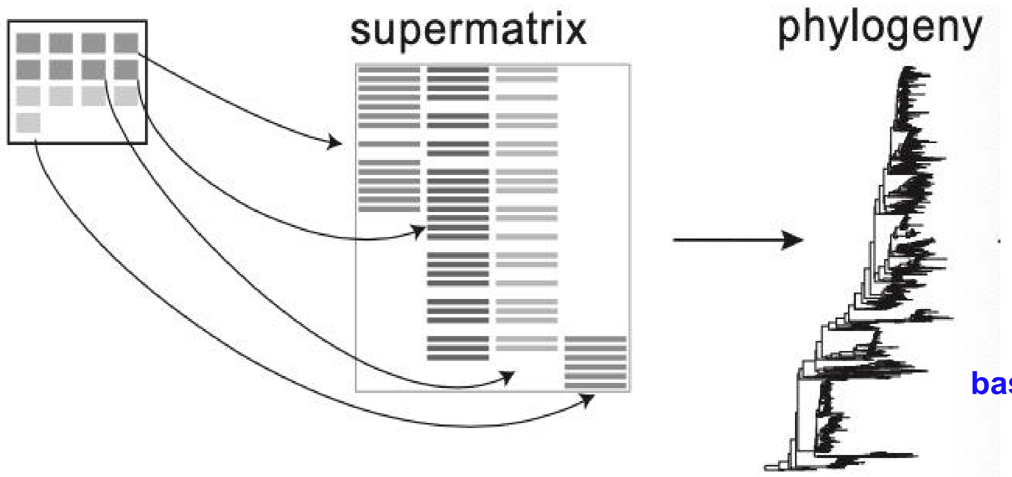
Royal Botanic Gardens Kew | Plants of the World Online

PyPHLAWD

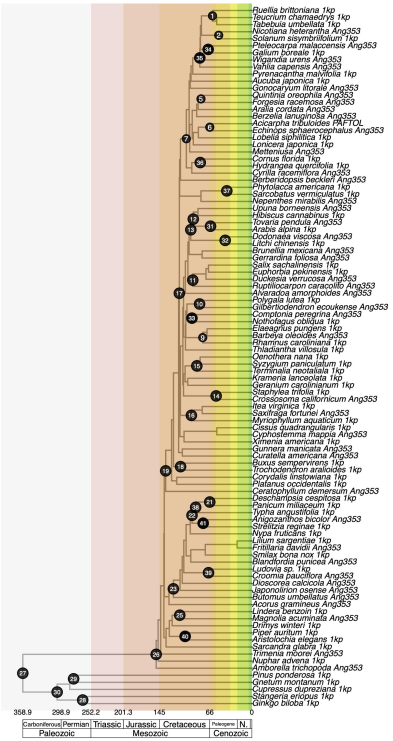
NCBI

PyPHLAWD

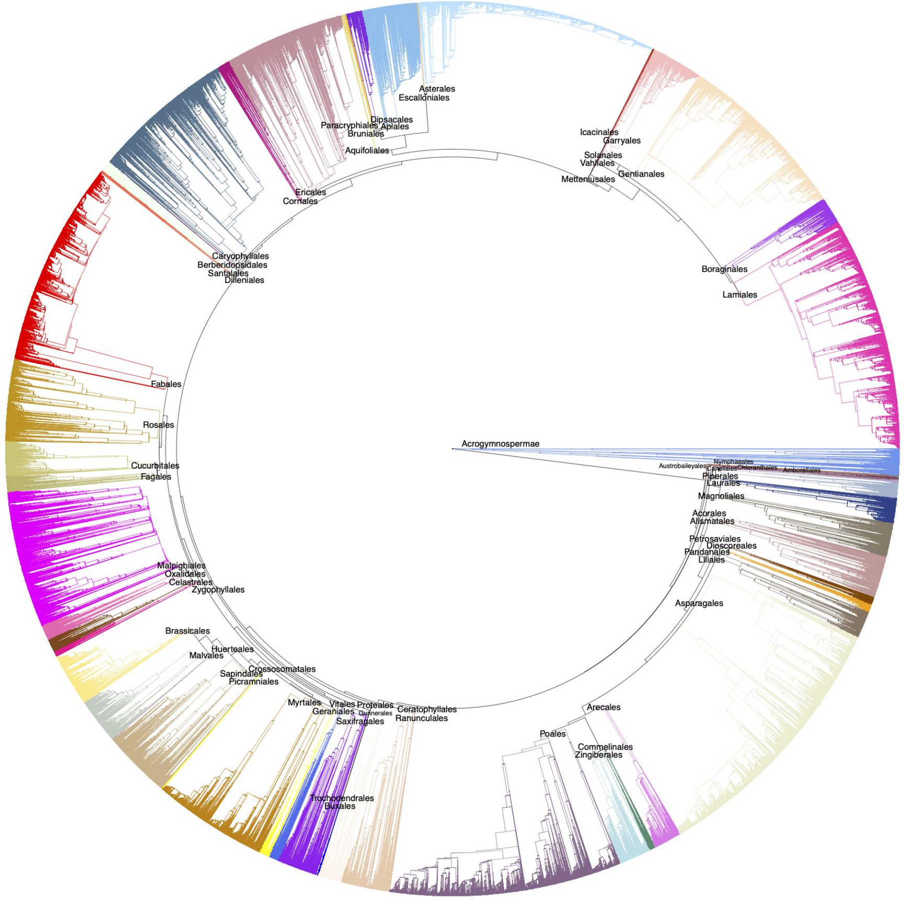
WCVP NCBI



Phylogenomics (1KP, PAFTOL)



phylogenomic backbone
based on genomes and transcriptomes



3,646,220-taxon tree of life of seed plant

Sun et al. (unpublished)

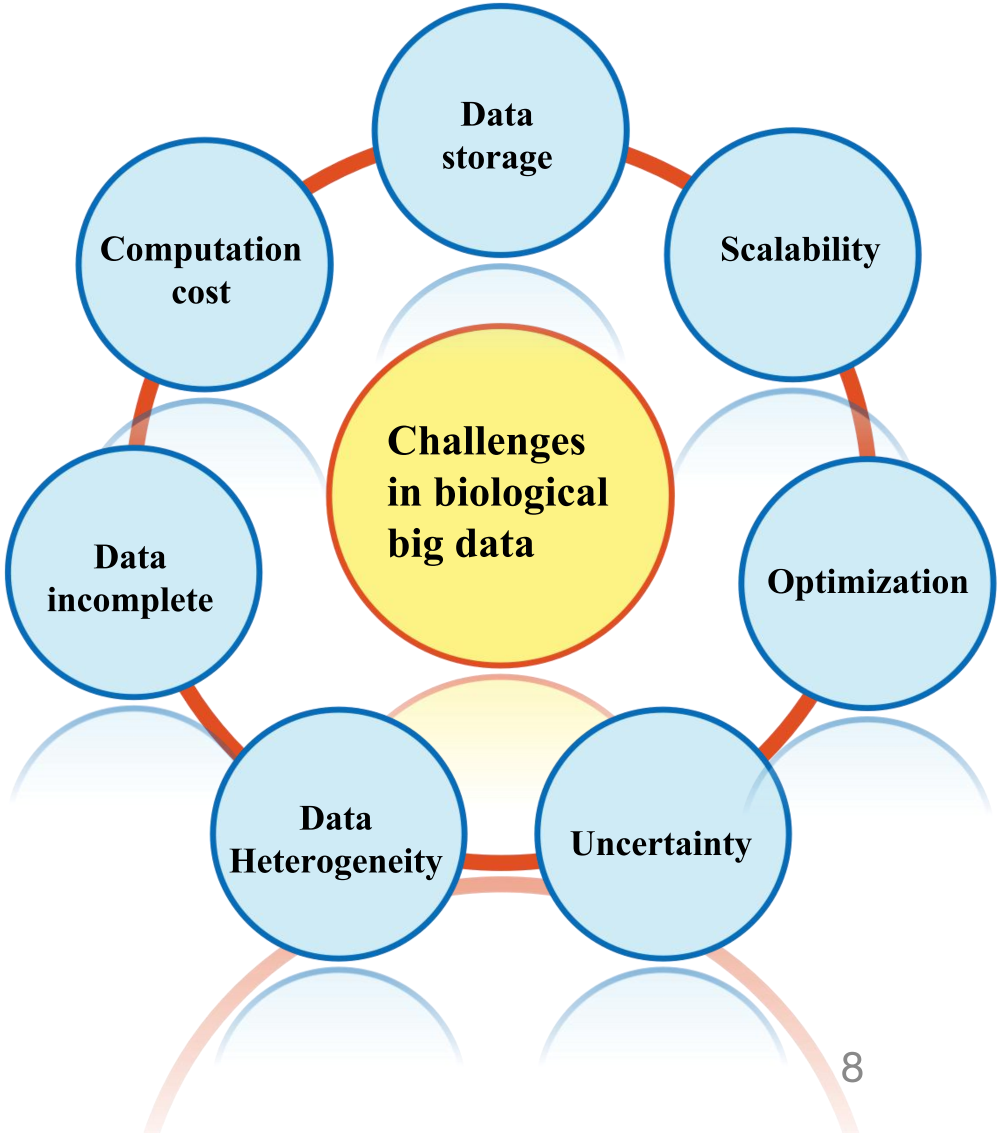
2. Biological Big Data

Big data is often characterised by the 3Vs:

- the large **volume** (数据体量)
- the wide **variety** (数据类型)
- the **velocity** (产生速度)

Biological big data cater to the needs of biological research, including -omics; traits, spatial data, species names, etc NLP related.

Small data is for people; Big data is for machines.



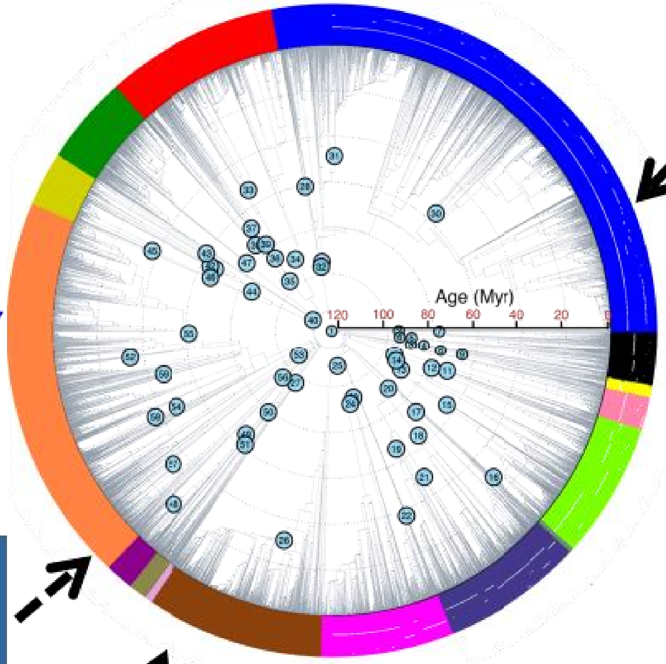
3. Applications of Tree of Life

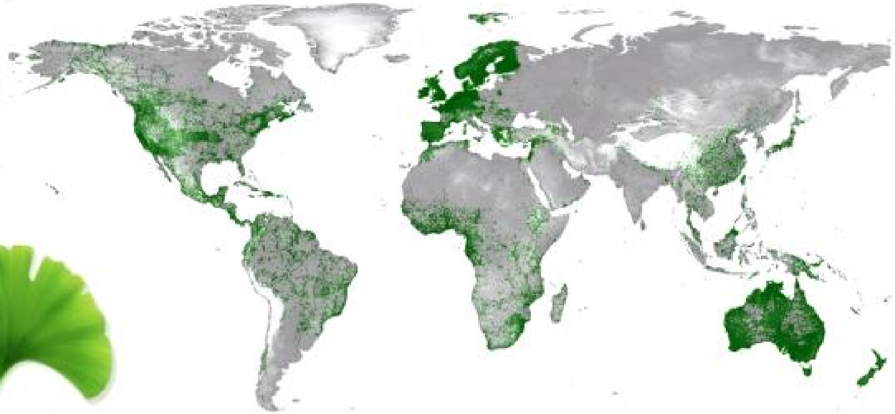
HiPerGator RESEARCH COMPUTING

AARHUS UNIVERSITY

GENOMEDK

NCBI





CVH

WorldClim - Global Climate Data

Free climate data for ecological modeling and GIS

IUCN RED LIST

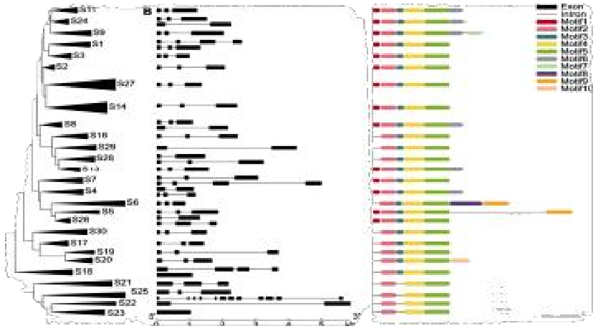
iDigBio Integrated Digitized Biocollections

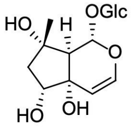
TRY

BIEN

UNIQUE forestry and land use

PhotosyntheticPathway Respiration LeafArea SLA GrowthForm PhenologyType LeafN LeafP MaxPlantHeight SeedMass

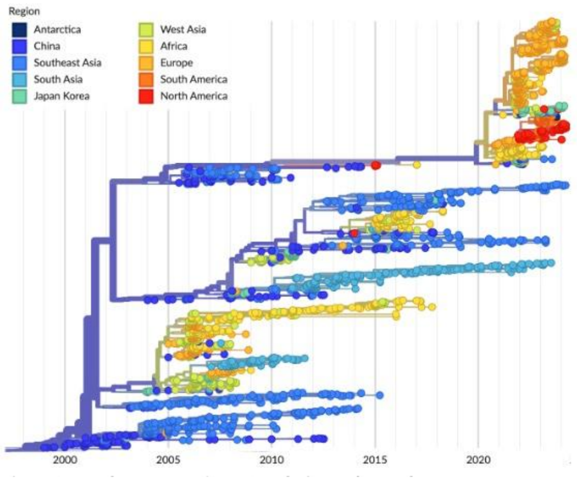


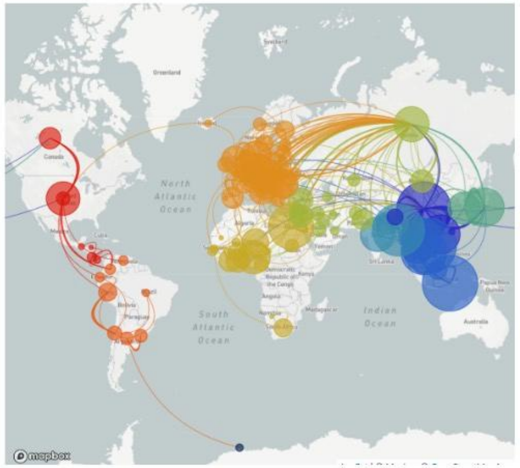


Tree of Life is a Platform

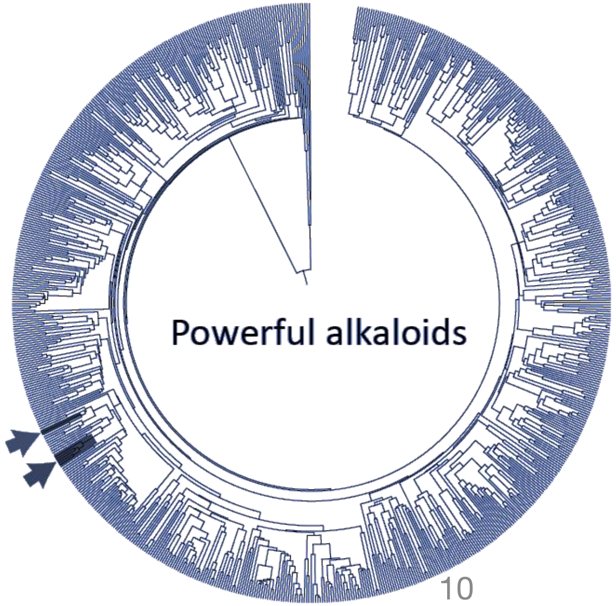
3. Applications of Tree of Life

- Conservation and Biodiversity
- Climate change
- Ecosystem
- Crop improvement
- Disease
- Drug discovery





Medicinal Hot Spots



3–1. Reticulation Cause Phylogenetic Discordance

ELSEVIER
Molecular Phylogenetics and Evolution
Volume 83, February 2015, Pages 156-166

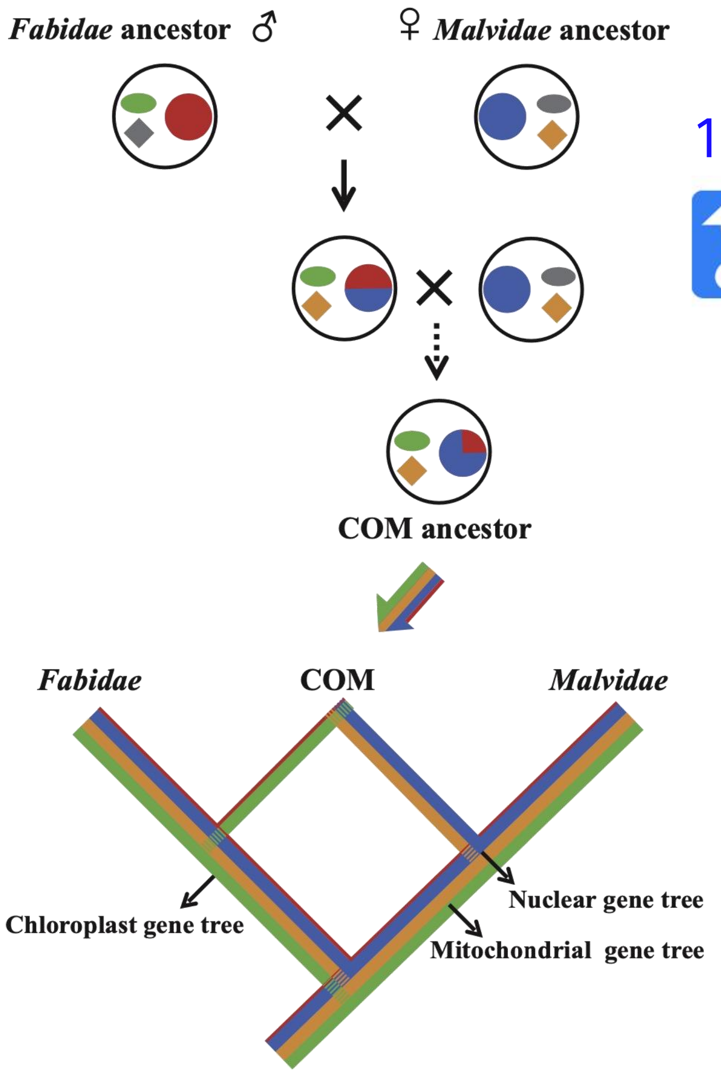
Deep phylogenetic incongruence in the angiosperm clade *Rosidae*

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]

169

Chloroplast, nuclear and mitochondrial genomic data, jointly confirmed conflict placement of the Celastrales-Oxalidales – Malpighiales (COM) clade, and our analyses showing ancient hybridization occurred between the ancestral *Fabidae* and *Malvidae* lineages, along with chloroplastid intergression give rise to the COM clade.

Sun et al. (2015)



3-2. Global Diversification Pattern of Rosids

华中农业大学 HUAZHONG AGRICULTURAL UNIVERSITY

被子植物多样性格局成因

Check for updates

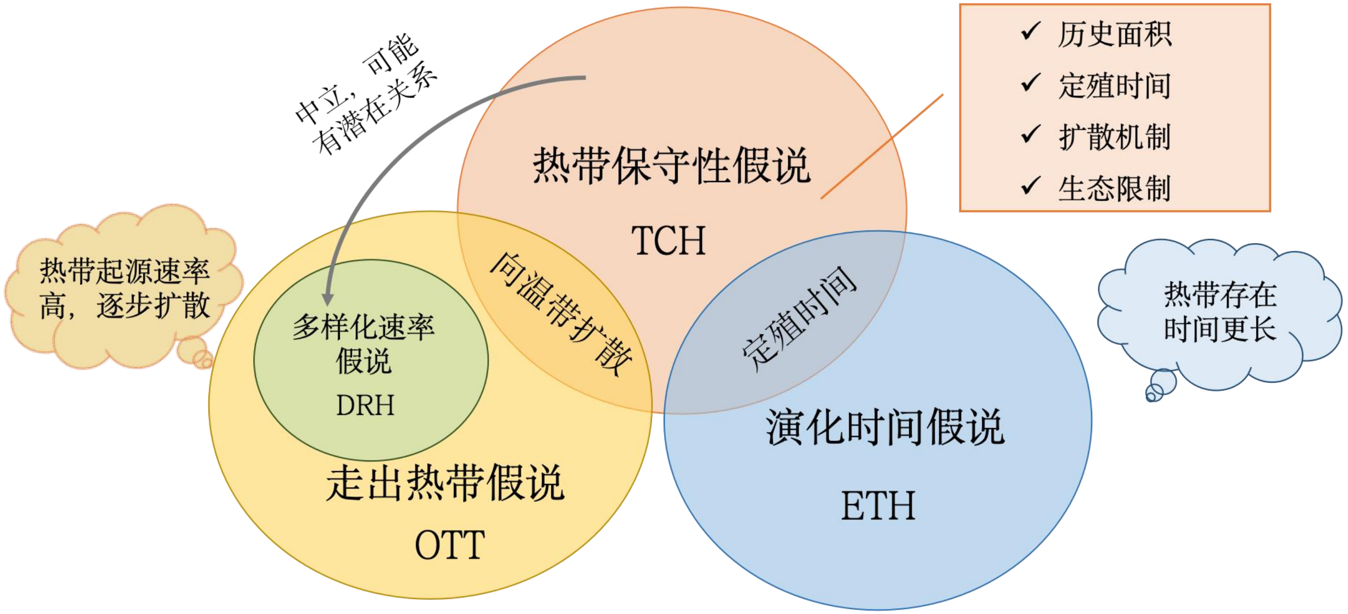
WILEY

SYNTHESIS

Journal of Biogeography

The Origins of the Latitudinal Diversity Gradient: Revisiting the Tropical Conservatism Hypothesis

John J. Wiens[1] | Michael J. Donoghue[1,2]



nature COMMUNICATIONS

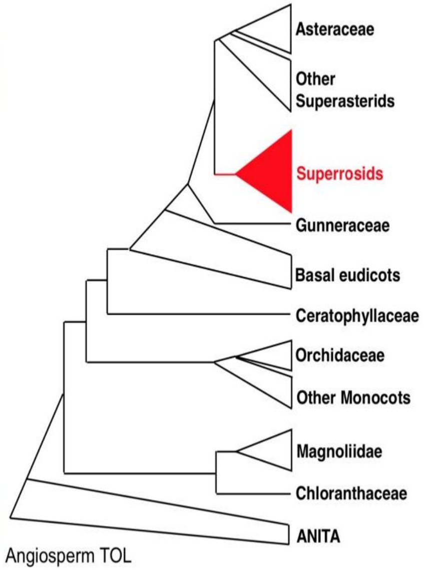
ARTICLE

Check for updates

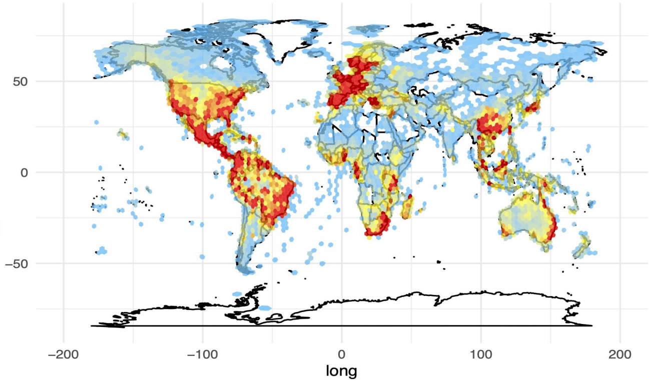
https://doi.org/10.1038/s41467-020-17116-5 OPEN

Recent accelerated diversification in rosids occurred outside the tropics

Miao Sun[1,2,3,8✉], Ryan A. Folk[4,8✉], Matthew A. Gitzendanner[5,6], Pamela S. Soltis[1,6,7], Zhiduan Chen[3], Douglas E. Soltis[1,5,6,7] & Robert P. Guralnick[1,6✉]



Magallón et al., 1999; Wang et al., 2009; Hinchliff et al., 2015; APG IV, 2016



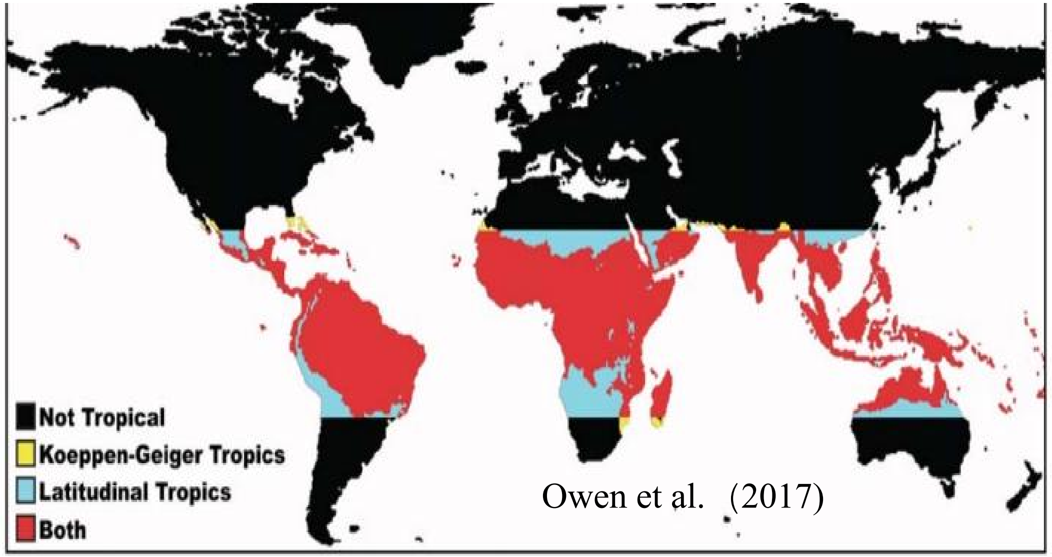
Sun et al. (2020)

3-2. Global Diversification Pattern of Rosids

華中農業大學 HUAZHONG AGRICULTURAL UNIVERSITY

Data Assembly

✓ **20,194-taxa dated rosid tree**
✓ Species distribution data
2,691,099 from iDigBio & GBIF
✓ Temperature Data Layers
35,164 historical oxygen isotope data ($\delta^{18}O$) (~113 Myr)
16,986 species AMT data from WorldClim (Geographic tropic traits)
Köppen-Geiger climatic tropic traits for 17,635 species



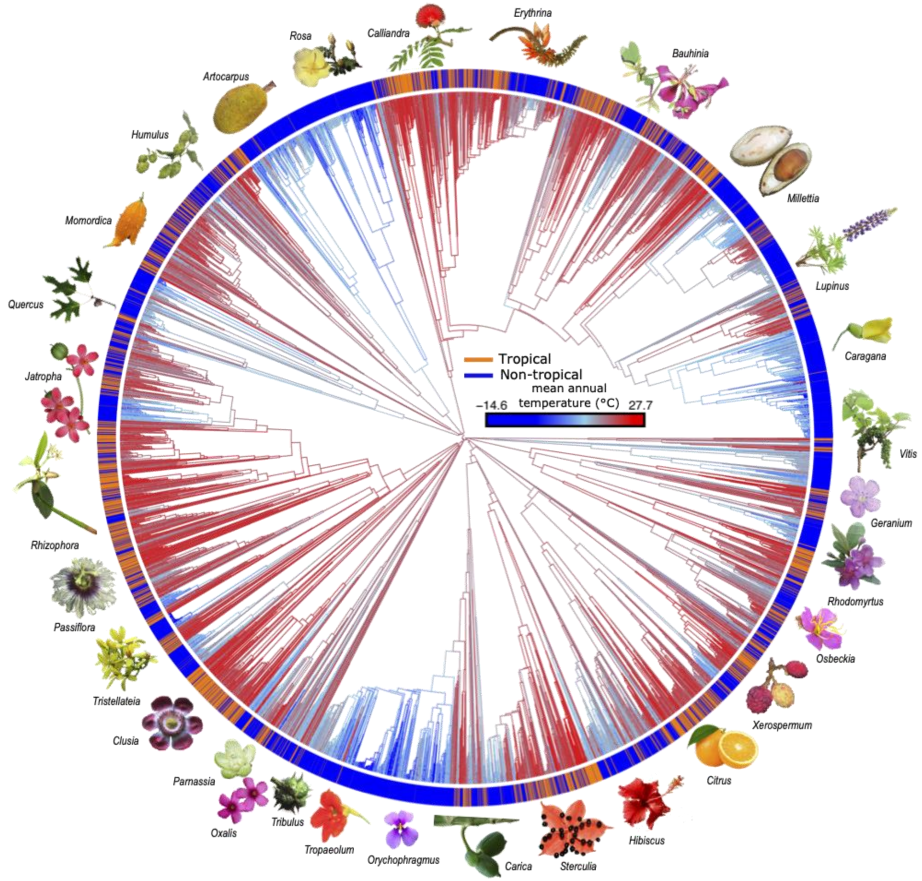
BAMM, RPANDA, DR, Pagel's λ tests, STRAPP, FiSSE, es-SIM and HiSSE comparative methods, spatical and uncertiany analyses

https://github.com/Cactusolo/rosid_NCOMMS-19-37964-T

3–2. Global Diversification Pattern of Rosids

Table 1 Summary table of semi-parametric tests for correlation between diversification rate and continuous mean annual temperature (es-SIM).

	es-SIM		
Order	**Rho (ρ)**	***p*-value**	**pv.adjust**
Brassicales	−0.3945	0.0180	0.2699
Celastrales	0.0185	0.9444	0.9444
Crossosomatales	−0.1255	0.8011	0.9444
Cucurbitales	−0.0471	0.9290	0.9444
Fabales	−0.3290	0.0060	0.0960
Fagales	−0.0767	0.4985	0.9444
Geraniales	0.0655	0.8117	0.9444
Huerteales	−0.6484	0.2413	0.9444
Malpighiales	−0.3256	0.0007	**0.0113**
Malvales	−0.1251	0.4392	0.9444
Myrtales	−0.2129	0.1113	0.9444
Oxalidales	−0.1205	0.5252	0.9444
Picramniales	−0.3169	0.7711	0.9444
Rosales	−0.2072	0.2186	0.9444
Sapindales	−0.0668	0.5012	0.9444
Vitales	−0.3762	0.1619	0.9444
Zygophyllales	−0.0581	0.7804	0.9444
Total tree	−0.2919	**0.0007**	/



Tip rates vs. MAT correlations significantly negative across entire tree

Sun et al. (2020)

3-2. Global Diversification Pattern of Rosids

Plot net diversification rate curves of major rosid clades against Tm

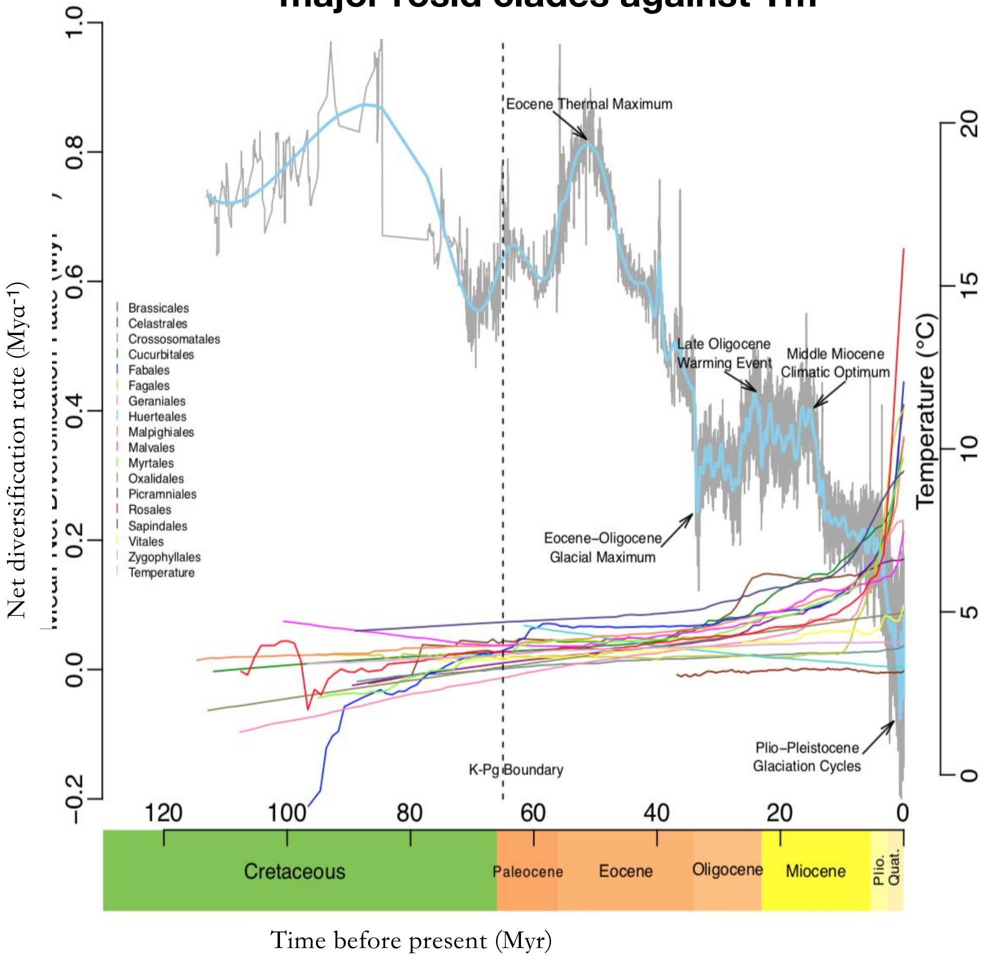
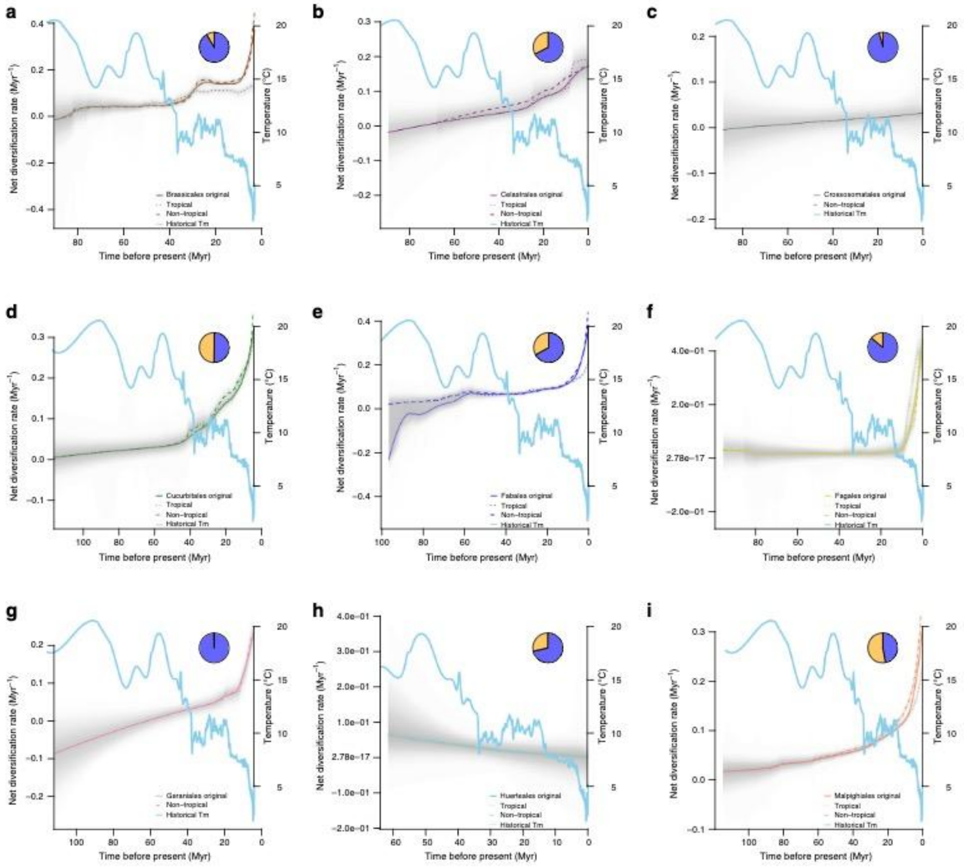


Table 2 Summary of correlation tests between tip rates and the two binary tropicality data sets in FiSSE.

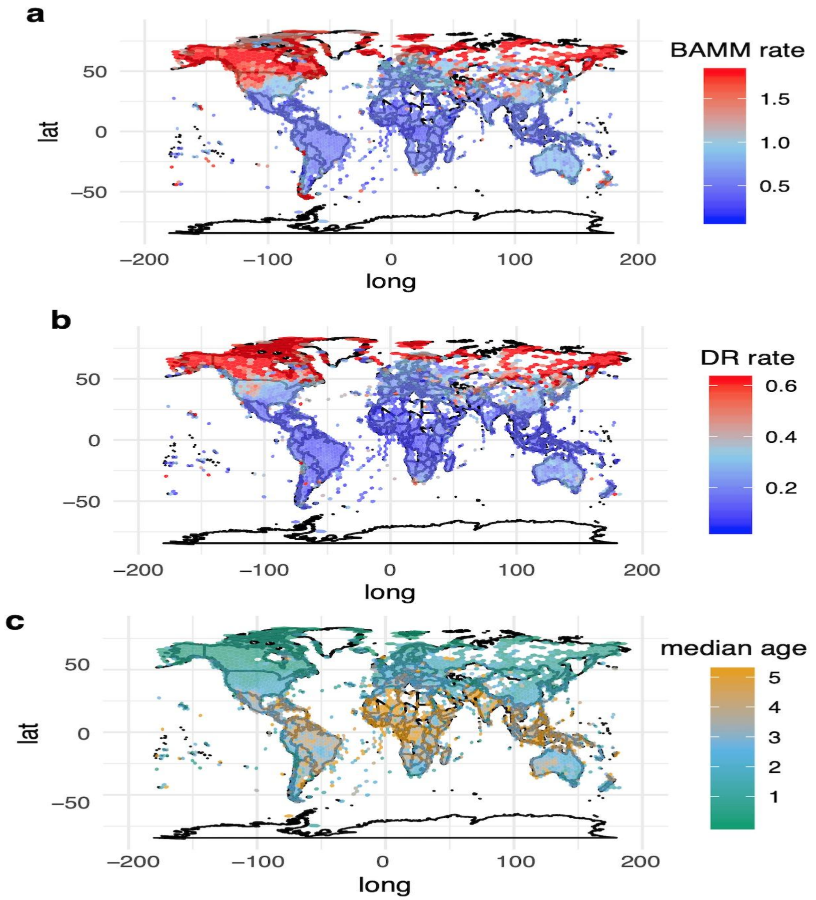
	Köppen-Geiger				Geographic			
Order	$\lambda_{non\text{-}tropical}$	$\lambda_{tropical}$	***p*-value**	**pv.adjust**	$\lambda_{non\text{-}tropical}$	$\lambda_{tropical}$	***p*-value**	**pv.adjust**
Brassicales	0.5198	0.1594	0.0985	0.5554	0.5079	0.4410	0.4865	0.7682
Celastrales	0.2208	0.2015	NA	NA	0.2309	0.2046	NA	NA
Crossosomatales	0.0776	0.1073	0.6104	0.6104	0.0893	0.0477	0.2248	0.7682
Cucurbitales	0.3033	0.2353	0.2168	0.5554	0.2919	0.2843	0.4945	0.7682
Fabales	0.6197	0.2946	0.0020	**0.0300**	0.6962	0.3450	0.0080	0.0959
Fagales	0.6032	0.3385	0.0320	0.4156	0.6276	0.4008	0.0609	0.6703
Geraniales	0.2919	0.0544	0.1578	0.5554	0.2988	0.2366	0.2378	0.7682
Huerteales	0.0278	0.0195	0.1718	0.5554	0.0278	0.0195	0.1768	0.7682
Malpighiales	0.4105	0.2312	0.0230	0.3217	0.4747	0.2582	0.0060	0.0779
Malvales	0.4066	0.2963	0.1898	0.5554	0.4305	0.3639	0.3666	0.7682
Myrtales	0.4766	0.2594	0.0440	0.4835	0.4972	0.3544	0.3437	0.7682
Oxalidales	0.1532	0.0854	0.1698	0.5554	0.1815	0.1503	NA	NA
Picramniales	0.0875	0.0488	0.1389	0.5554	NA	NA	NA	NA
Rosales	0.6407	0.3218	0.0390	0.4675	0.6795	0.4339	0.1149	0.7682
Sapindales	0.2576	0.2202	0.1159	0.5554	0.2815	0.2280	0.1089	0.7682
Vitales	0.7599	0.1055	NA	NA	0.9223	0.1014	NA	NA
Zygophyllales	0.0628	0.0537	0.2777	0.5554	0.0570	0.0663	0.7682	0.7682
Total tree	0.5047	0.2552	**0**	/	0.5552	0.3089	**0**	/

- ✓ **Tip rates vs. both geographic/climatic tropicality traits significantly correlated across entire tree**
- ✓ **$\lambda_{non\text{-}tropical}$ ~2 times $\lambda_{tropical}$**

3-2. Global Diversification Pattern of Rosids



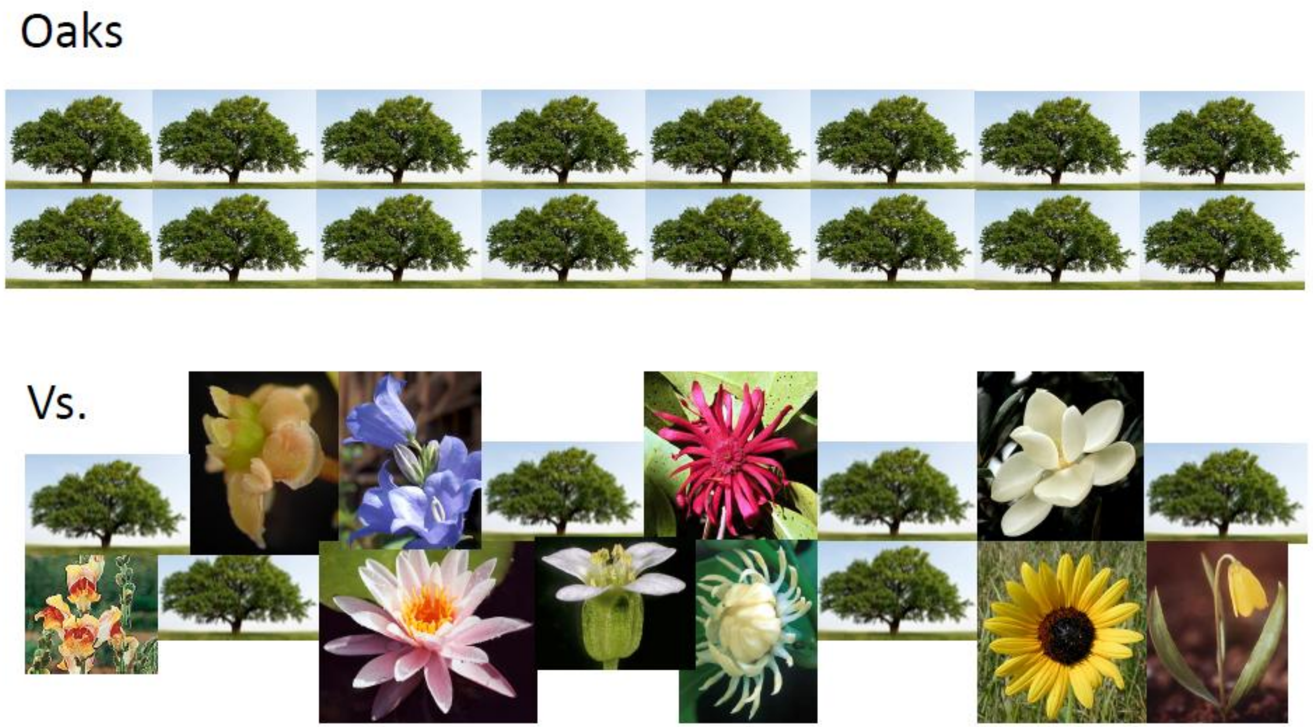
Sun et al. (2020)

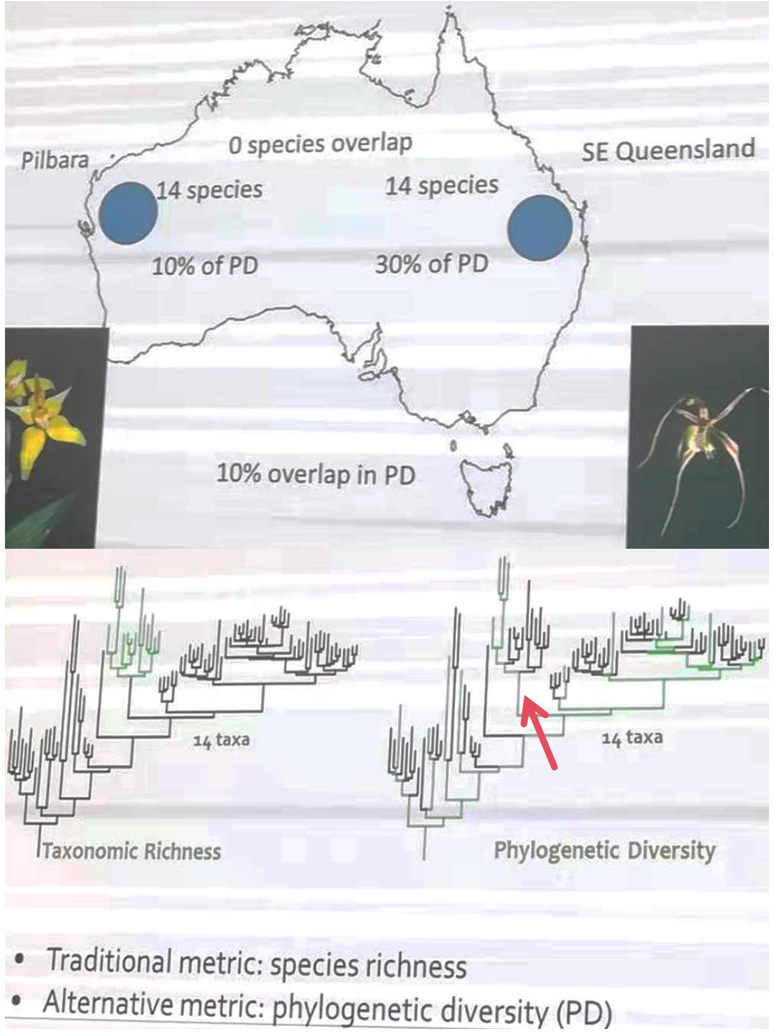


Negative relationship; Increased from 15 Myr, global cooling; Tropical species age older, rates lower vs. those outside tropics.

3–3. Phylogenetic Diversity & Conservation

Species richness vs.
Phylogenetic diversity (PD)





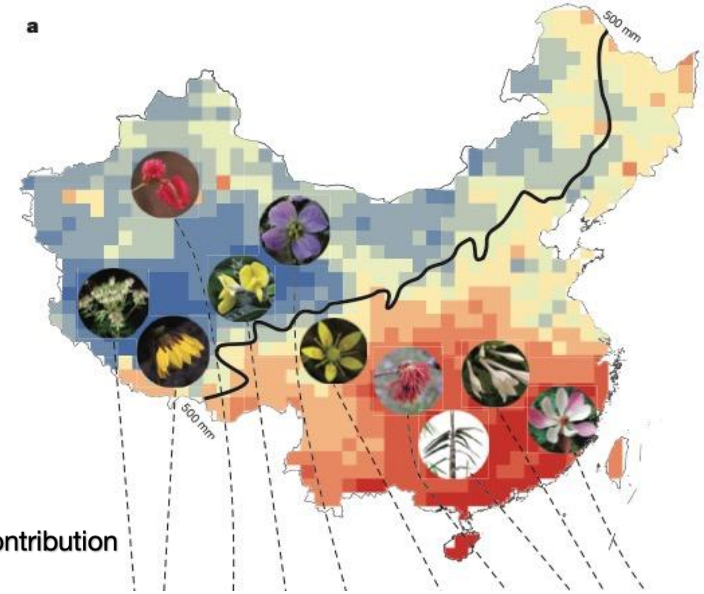
(Courtesy from Dr. Joe Miller)

3-3. Phylogenetic Diversity & Conservation

nature
International journal of science

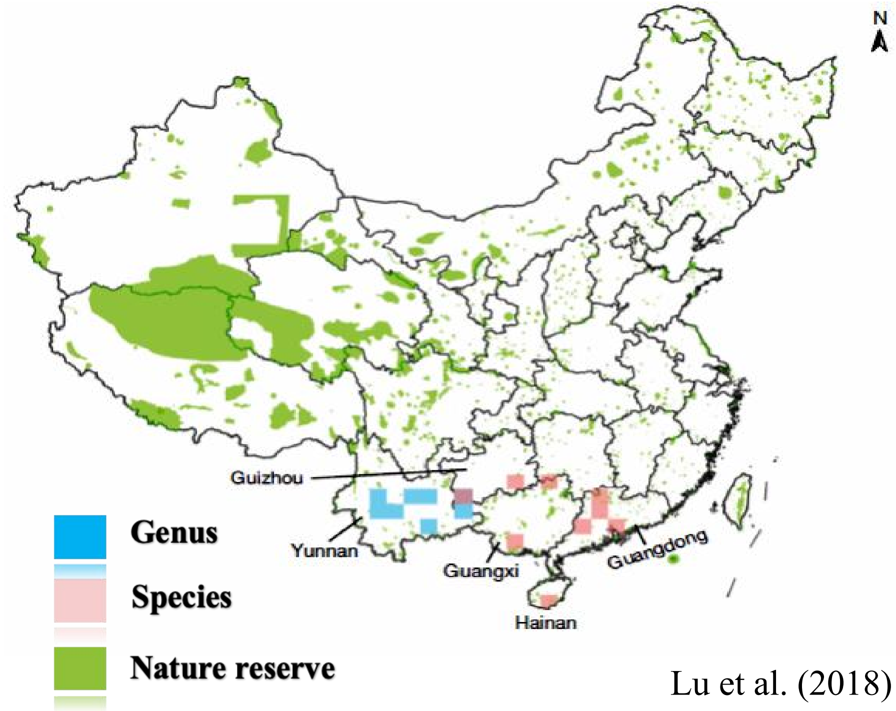
Evolutionary history of the angiosperm flora of China

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 ✉



*equal contribution

Build the first Chinese
26,978-taxa Angiosperm phylogeny



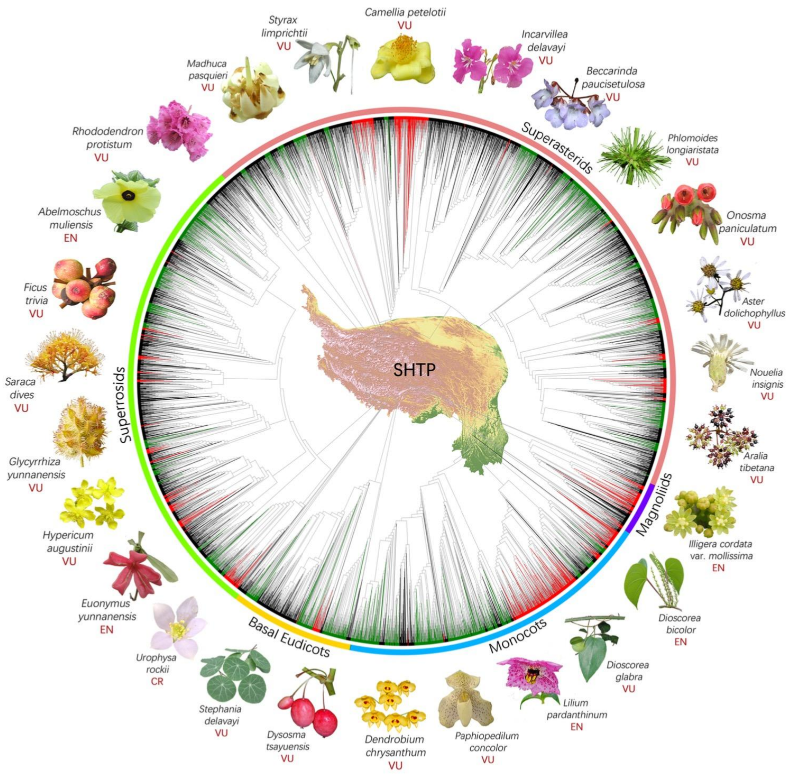
Lu et al. (2018)

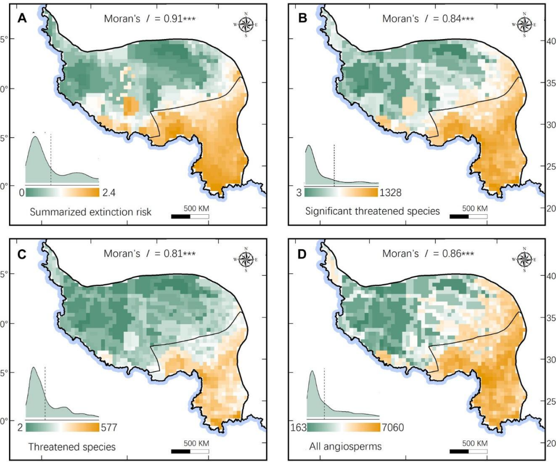
Provide suggestions for the conservation and future nature reserve plans and strategy

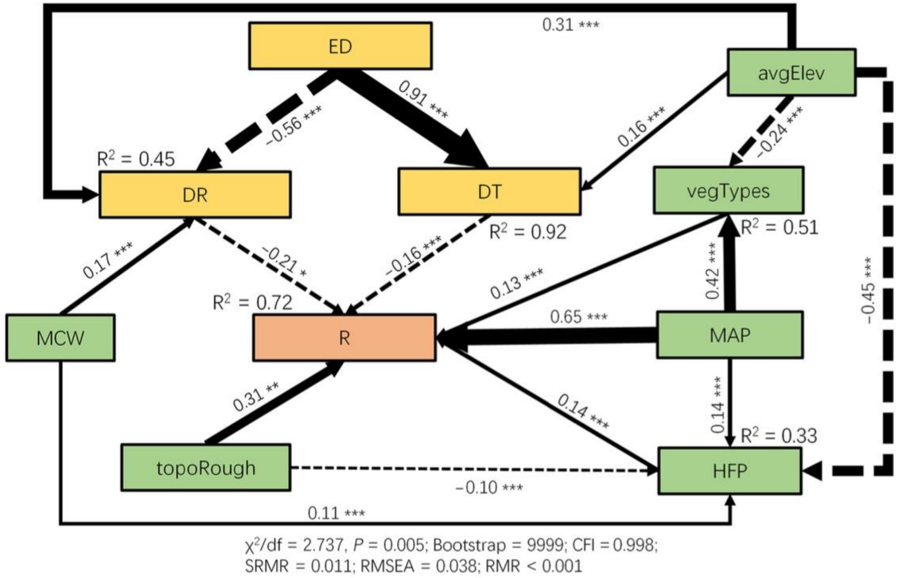
3–4. Phylogeny, Extinction Risk & Conservation

Flowering plants at Sino-Himalaya and the Tibetan Plateau face increased extinction risk

Li-Na Zhao[1,2,3†], Yun Liu[4†], Jian-Fei Ye[5,6], Bing Liu[1,2], Hai-Hua Hu[1,2], Li-Min Lu[1,2], Jiang Chang[7], Robert P. Guralnick[8], Miao Sun[9,10*], and Zhi-Duan Chen[1,2,11*]







Identified >3K species threatened not in Red List
Nonrandon, species-poor more likely to extinction
High risks in mountainous areas of southwest China
Evolutionary/environmental impact on extinction risk

Zhao et al. (2025)

3–5. Other Evolutionary Pattern Discovery

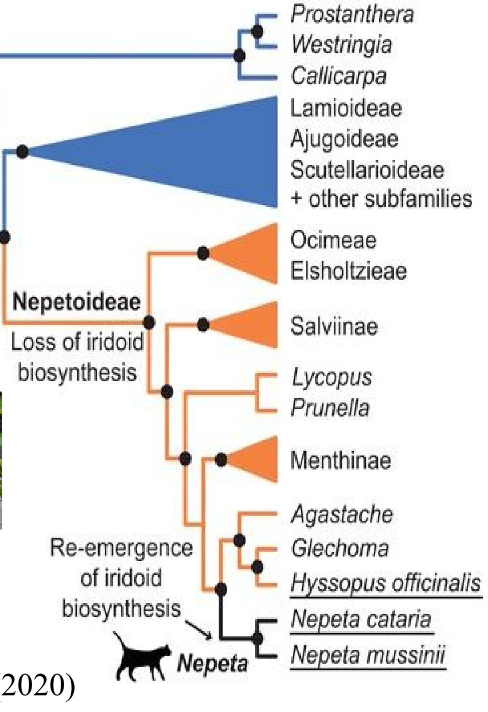
HUAZHONG AGRICULTURAL UNIVERSITY

SCIENCE ADVANCES | RESEARCH ARTICLE

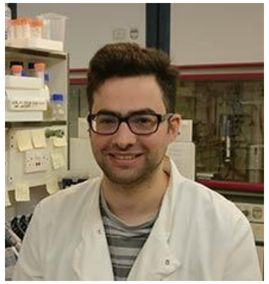
BIOSYNTHESIS

The evolutionary origins of the cat attractant nepetalactone in catnip

Benjamin R. Lichman[1]*, Grant T. Godden[2], John P. Hamilton[3], Lira Palmer[4], Mohamed O. Kamileen[4], Dongyan Zhao[3†], Brieanne Vaillancourt[3], Joshua C. Wood[3], Miao Sun[2], Taliesin J. Kinser[2,5], Laura K. Henry[6], Carlos Rodriguez-Lopez[4], Natalia Dudareva[6,7,8], Douglas E. Soltis[2,5], Pamela S. Soltis[2], C. Robin Buell[3,9,10]*, Sarah E. O'Connor[4]*



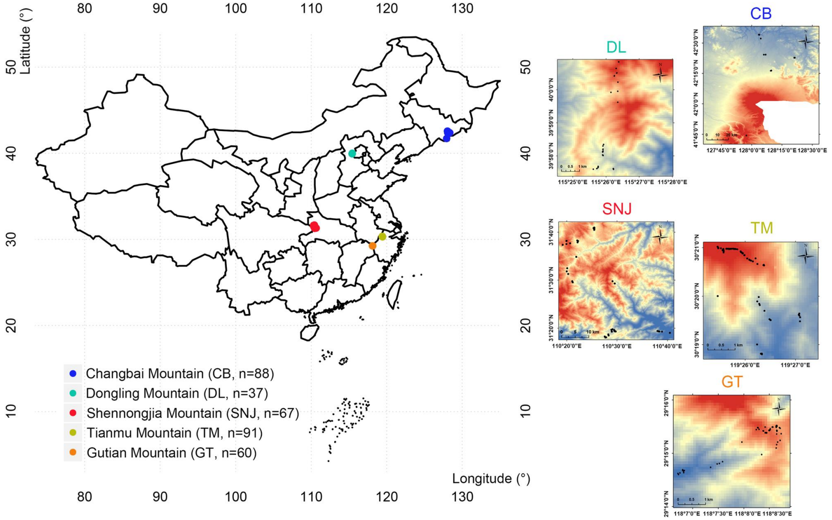
catmint



Dr. Benjamin R. Lichman
Plant Biology CNAP,
Department of Biology,
University of York

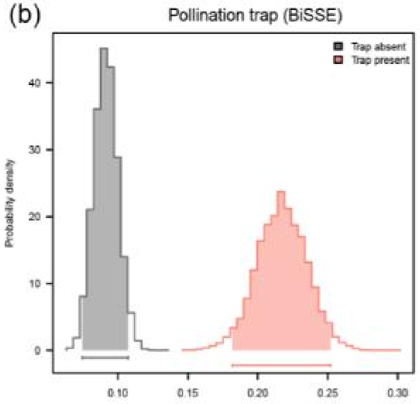
Lichman et al. (2020)

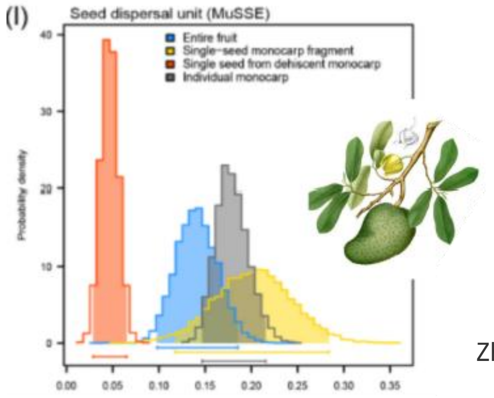
plants — soil fungi   Yang et al. (2018)

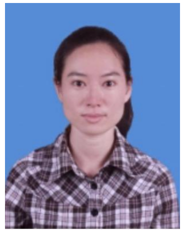


Dr. Haiyan Chu
Institute of Soil
Science, Chinese
Academy of
Sciences

traits — rates   Xue et al. (2019)







Dr. Bine Xue
College of
Horticulture,
Zhongkai University

The 3rd Party Evaluation

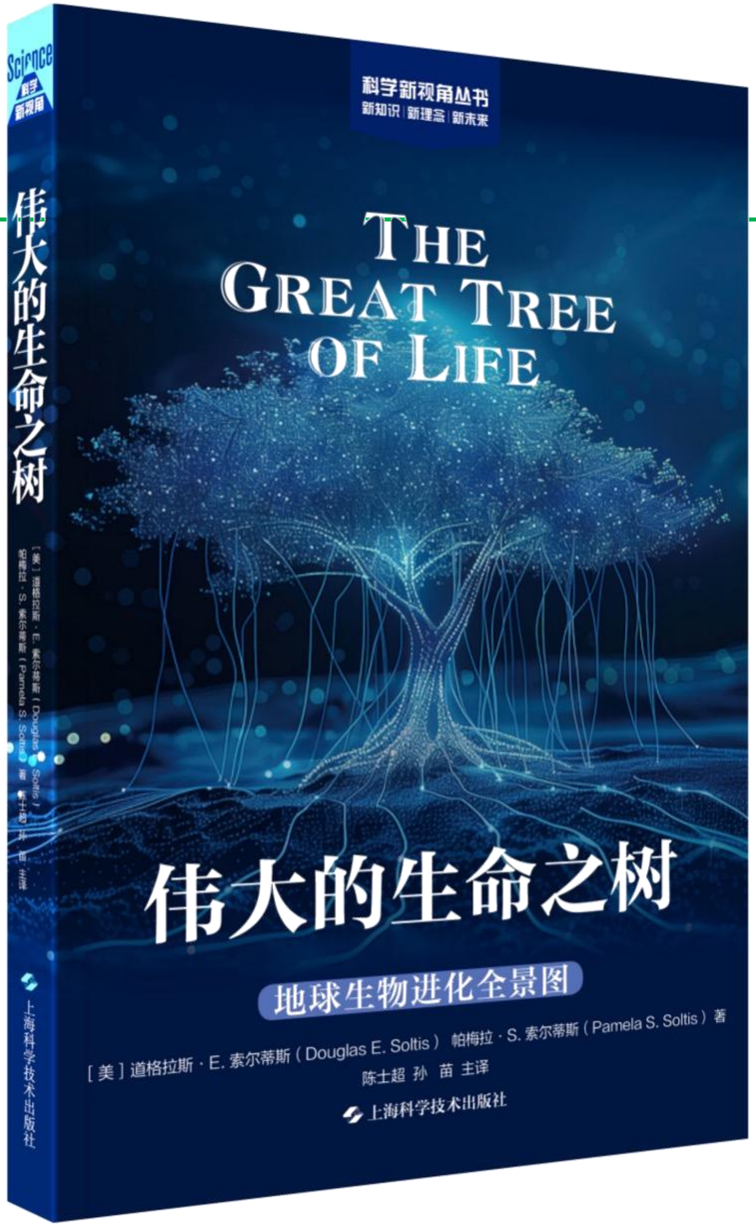
华中农业大学 HUAZHONG AGRICULTURAL UNIVERSITY

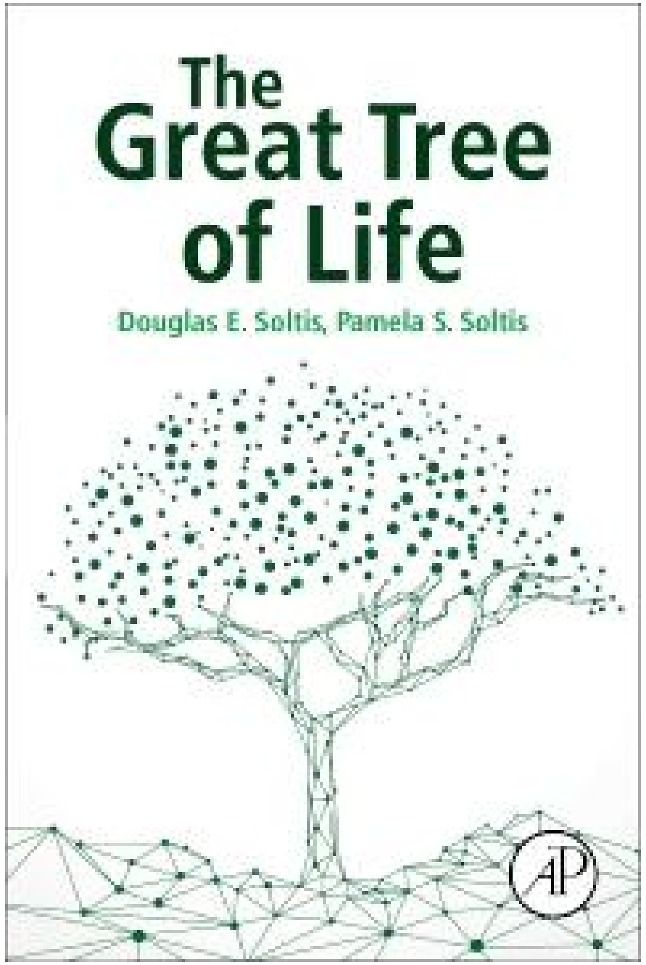
蔷薇类的工作分别得到美国科学院院士 Michael J. Donoghue 和 Doug E. Soltis 的肯定和引用

Summary

- **The Tree of life** is a "Family Tree" of all life, and an inventory of biodiversity
- **The Tree of Life** reveals evolutionary history (the past and present) of extent species
- **The Tree of Life** unveils how plants respond to climate change
- **The Tree of Life** plays an important role of measure bidoversity and conservation
- **The Tree of Life** is essential for biochemical compounds, and key traits discovery for medical drugs, and horticultural and crop improvment

... and in many other ways...



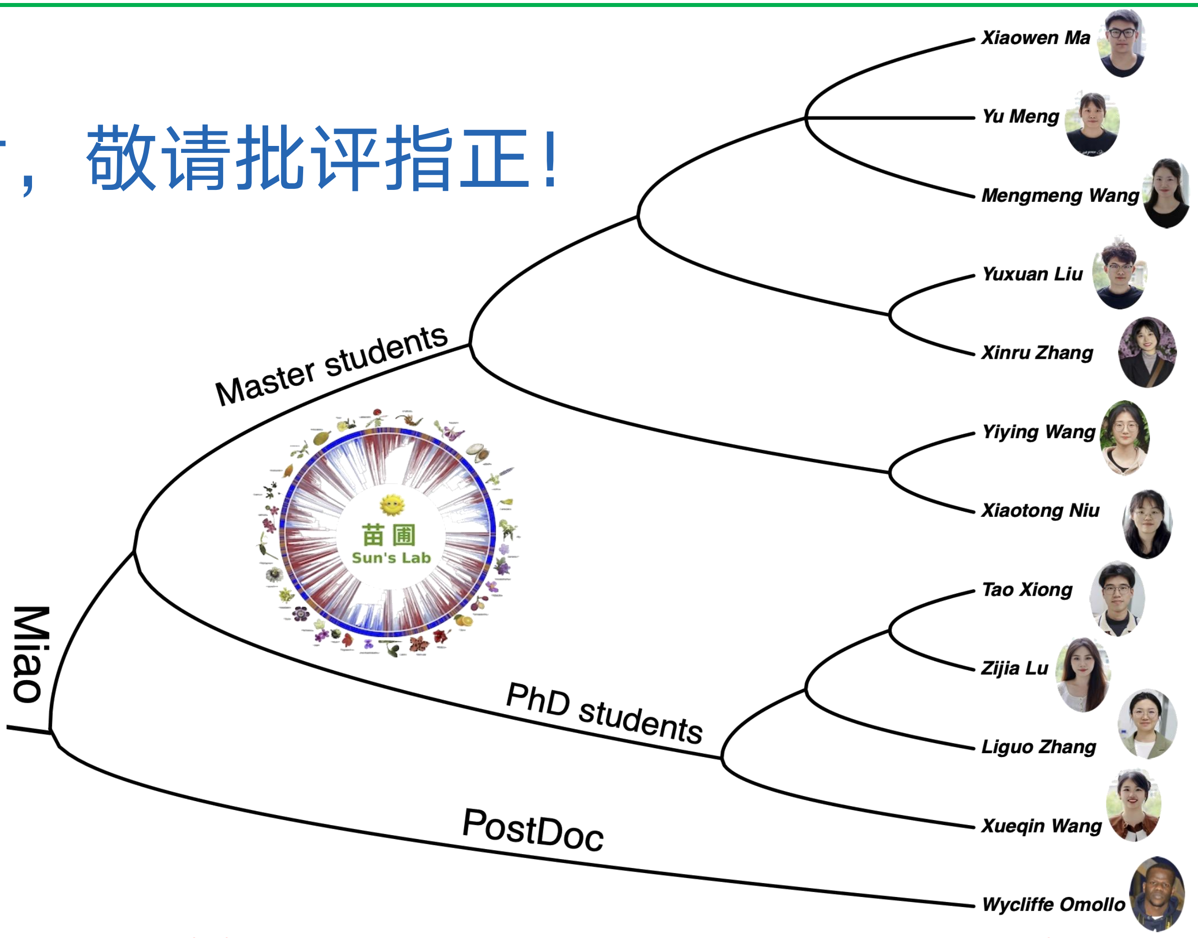


Soltis & Soltis, 2019

Acknowledgements

华中农业大学
HUAZHONG AGRICULTURAL UNIVERSITY

谢谢，敬请批评指正！



Doug Soltis
Pam Soltis
Ryan A. Folk
Sangtae Kim
Hiro Shimai
Imalka M. Kahandawala
Alexey P. Seregin
Yury O. Kopylov-Guskov
Chien-Hsun Huang
Zhiduan Chen
Qingfen Wang
Zhun Xu
Jiaxin Yang
Tuo Yang
XinXin Fu
Zhuzhu Yang

Guangwan Hu
Shengdan Wu
Zhuoming Shen
Qiang Fu
Kailing Dong
Hanning Duan
Dongxian Xu
Haicheng Zhou
Jianfei Ye
Bing Liu
Shuangwen Deng
Chunlei Xiang
Xiangqin Yu

We are recruiting graduate students and Post-docs!