

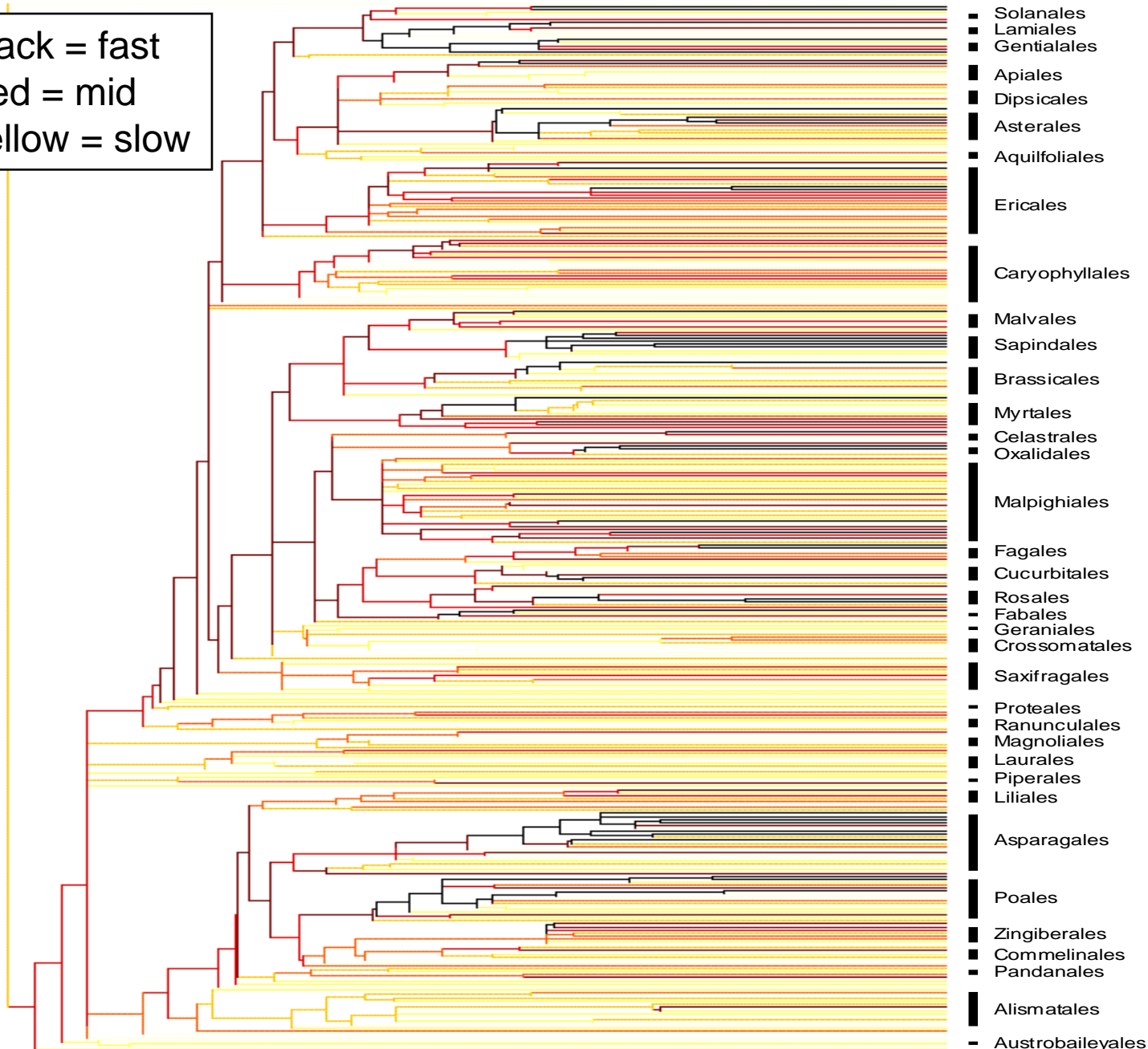
# Work Package 3: construction a phylogenetic hypothesis of all tree genera in French Guiana

... on behalf of Vincent  
Savolainen

# Background: APG2 phylogeny

- An international effort led by Mark Chase / Kew gardens & Doug & Pam Soltis / Univ Florida
  - the Angiosperm Phylogeny Group (1998, 2003)
- Angiosperm classification has been revised on the basis of molecular analyses :
  - Paleoherbs, Monocots, Eudicots etc
  - 462 well-supported families
- Calibration of this phylogenetic hypothesis
  - using fossil data as calibration points
  - and based on statistical techniques that make all tips of the tree equidistant to the root

Black = fast  
Red = mid  
Yellow = slow



# Towards APG3

- Aim : to obtain a phylogenetic hypothesis for as many genera as possible

**2003 Mayr Award Winner**

*Syst. Biol.* 54(2):183–196, 2005  
Copyright © Society of Systematic Biologists  
ISSN: 1063-5157 print / 1076-836X online  
DOI: 10.1080/10635150590923254

## **Towards Building the Tree of Life: A Simulation Study for All Angiosperm Genera**

NICOLAS SALAMIN,<sup>1,2</sup> TREVOR R. HODKINSON,<sup>1</sup> AND VINCENT SAVOLAINEN<sup>2</sup>

<sup>1</sup>*Department of Botany, University of Dublin, Trinity College, Dublin 2, Ireland; E-mail: nicolas.salamin@unil.ch*

<sup>2</sup>*Molecular Systematics Section, Jodrell Laboratory, Royal Botanic Gardens, Kew, Richmond Surrey, TW9 3DS, London, UK*

**Abstract.**—Comprehensive phylogenetic trees are essential tools to better understand evolutionary processes. For many groups of organisms or projects aiming to build the Tree of Life, comprehensive phylogenetic analysis implies sampling hundreds to thousands of taxa. For the tree of all life this task rises to a highly conservative 13 million. Here, we assessed the performances of methods to reconstruct large trees using Monte Carlo simulations with parameters inferred from four large angiosperm DNA matrices, containing between 141 and 567 taxa. For each data set, parameters of the HKY85+Γ model were estimated and used to simulate 20 new matrices for sequence lengths from 100 to 10,000 base pairs. Maximum parsimony and neighbor joining were used to analyze each simulated matrix. In our simulations, accuracy was measured by counting the number of nodes in the model tree that were correctly inferred. The accuracy of the two methods increased very quickly with the addition of characters before reaching a plateau around 1000 nucleotides for any sizes of trees simulated. An increase in the number of taxa from 141 to 567 did not significantly decrease the accuracy of the methods used, despite the increase in the complexity of tree space. Moreover, the distribution of branch lengths rather than the rate of evolution was found to be the most important factor for accurately inferring these large trees. Finally, a tree containing 13,000 taxa was created to represent a hypothetical tree of all angiosperm genera and the efficiency of phylogenetic reconstructions was tested with simulated matrices containing an increasing number of nucleotides up to a maximum of 30,000. Even with such a large tree, our simulations suggested that simple heuristic searches were able to infer up to 80% of the nodes correctly. [Angiosperms; maximum parsimony; Monte Carlo simulations; neighbor joining; taxon sampling; Tree of Life.]

# Why ecologists became interested in phylogenies?

- Evolutionary relationships between organisms are key to an understanding of their ecological interactions (why species are so similar? or different? Where do they come from?)
- Ecological processes are key to understand the processes underlying species diversification (what drives speciation? extinction? Why species look the way they look?)

*Ecology*, 87(7) Supplement, 2006, pp. S163–S165  
© 2006 by the Ecological Society of America

## PHYLOGENETIC ECOLOGY AT WORLD SCALE, A NEW FUSION BETWEEN ECOLOGY AND EVOLUTION

MARK WESTOBY<sup>1</sup>

*Department of Biological Sciences, Macquarie University, Sydney, New South Wales 2109 Australia*

*Abstract.* One fusion between ecology and evolution is well established, under the title of population biology. The years 2006–2020 will see a new fusion, likely to prove equally creative. Inputs from ecology to this second fusion will be worldwide data sets for ecological traits across many species. Inputs from evolution will be phylogenetic trees with well-resolved topology and with increasingly tight geological dates for each branch point. There will be

Factoring out similarities due to  
evolutionary history  
to answer important ecological  
questions

# why do the tropics contain more species?



Amir Takhtajan

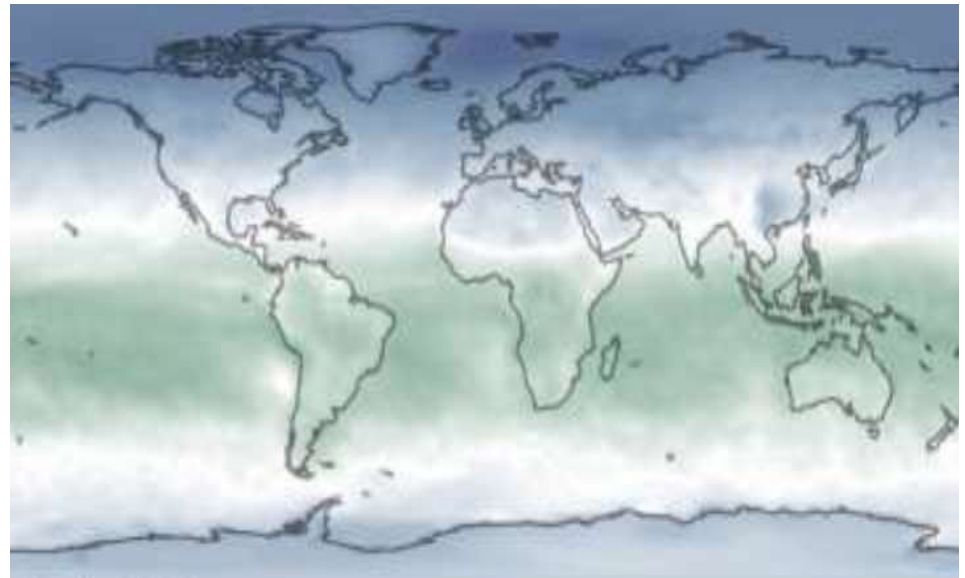
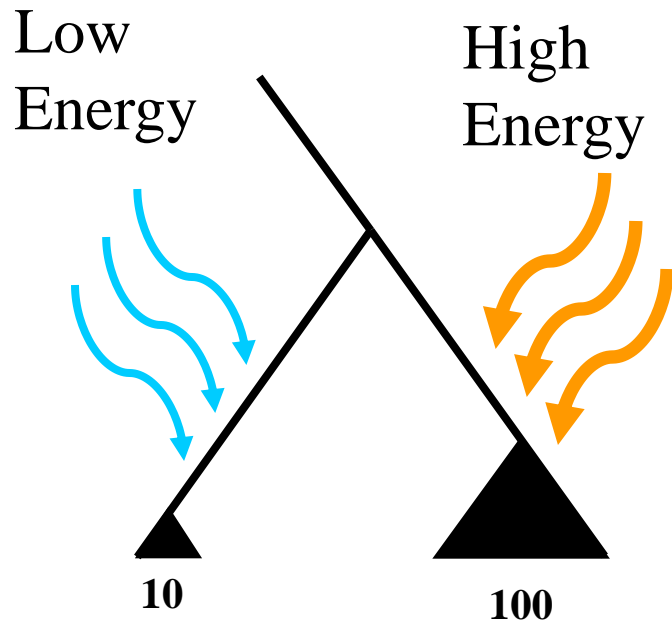
## SOLAR ULTRAVIOLET ON TROPICAL MOUNTAINS: CAN IT AFFECT PLANT SPECIATION?

The subalpine rain forests and true alpine zones (Grubb 1977) of high tropical mountains contain plant species with unusual morphologies, including cushion plants, microphyllous shrubs, and pachycaulous trees (Cuatrecasas 1968; Troll 1968). Although these communities are not species rich, they include a high percentage of endemics. Hedburg (1969) estimated that 80% of the East African alpine flora is endemic to those mountains. Smith (1975) listed 50 endemics (among 139 species) from the Irian Jaya Mountains of New Guinea, and most of the other species grow only on New Guinea mountains. Of the herbaceous subalpine plants he discussed from Mount Kinabalu, Borneo (Smith 1970), 40% grow only on that mountain. All of these tropical mountain floras are taxonomically more closely linked with temperate floras than with floras of the surrounding lowland forest, which serves as a medium of isolation rather than a source of species.

Lee, Am Nat (1980)

# The Faster Evolution Hypothesis

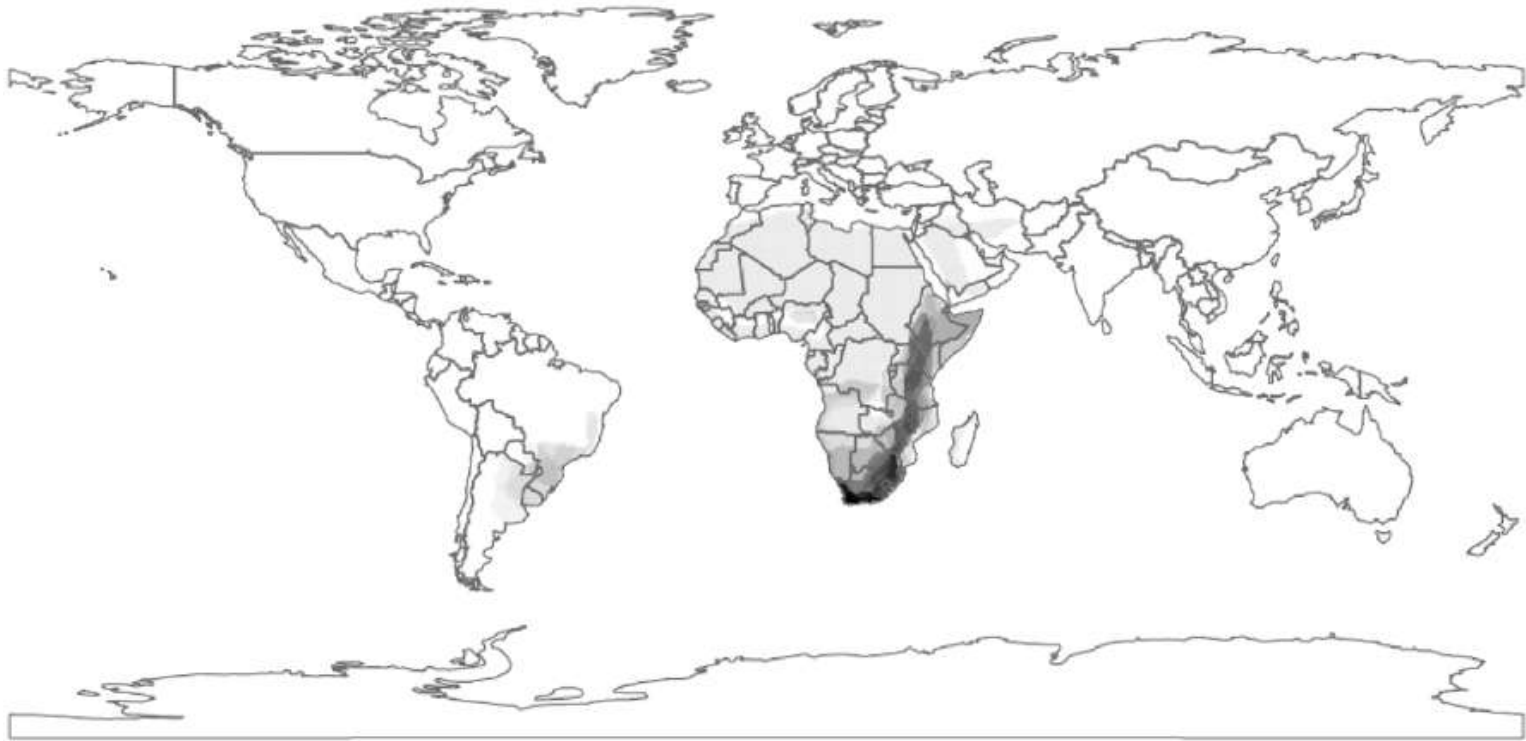
More radiation in the tropics => higher molecular evolution => more diversification



Net radiation (February 2002)



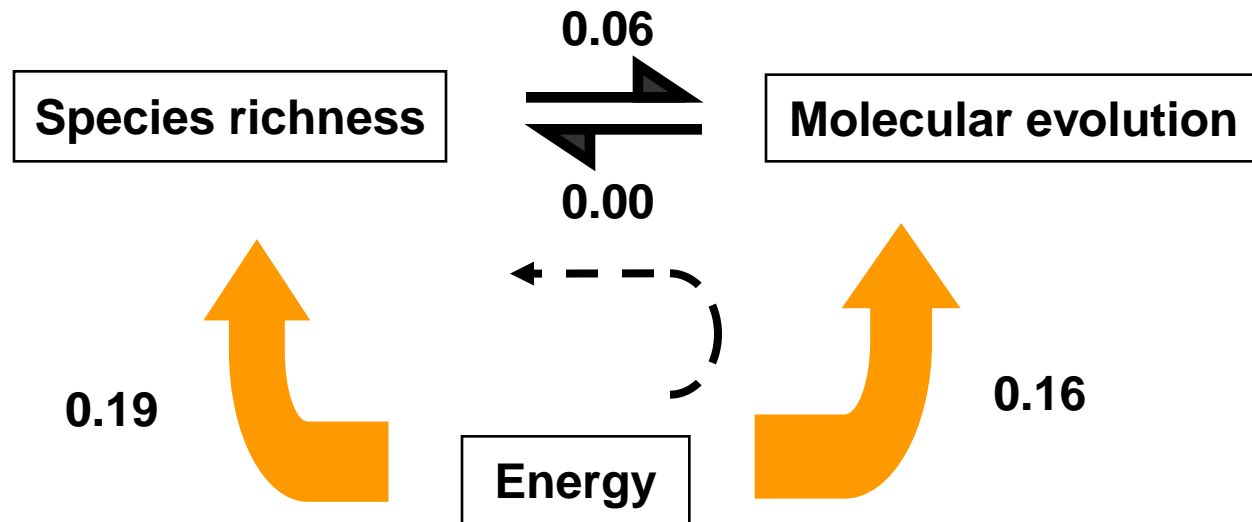
# A test with the Iridaceae



Cape and atlantic coast forest of Brazil have excess of *Iris* species after area, age and energy factored out

both faster molecular evolution and higher species richness in high energy environments,

BUT there is no support for the **faster evolution theory**



Adding both energy and molecular rates as predictors of species richness causes mol. rates to drop out of the model

Adding a timeframe:

dating events that mattered for  
the buildup of biodiversity

# Explosive Radiation of Malpighiales Supports a Mid-Cretaceous Origin of Modern Tropical Rain Forests

Charles C. Davis,<sup>1,\*</sup> Campbell O. Webb,<sup>2,\*</sup> Kenneth J. Wurdack,<sup>3,\*</sup> Carlos A. Jaramillo,<sup>4,5</sup> and Michael J. Donoghue<sup>2,3</sup>

1. Department of Ecology and Evolutionary Biology, University of Michigan Herbarium, Ann Arbor, Michigan 48108-2287;

2. Department of Ecology and Evolutionary Biology, Yale University, P.O. Box 208106, New Haven, Connecticut 06520;

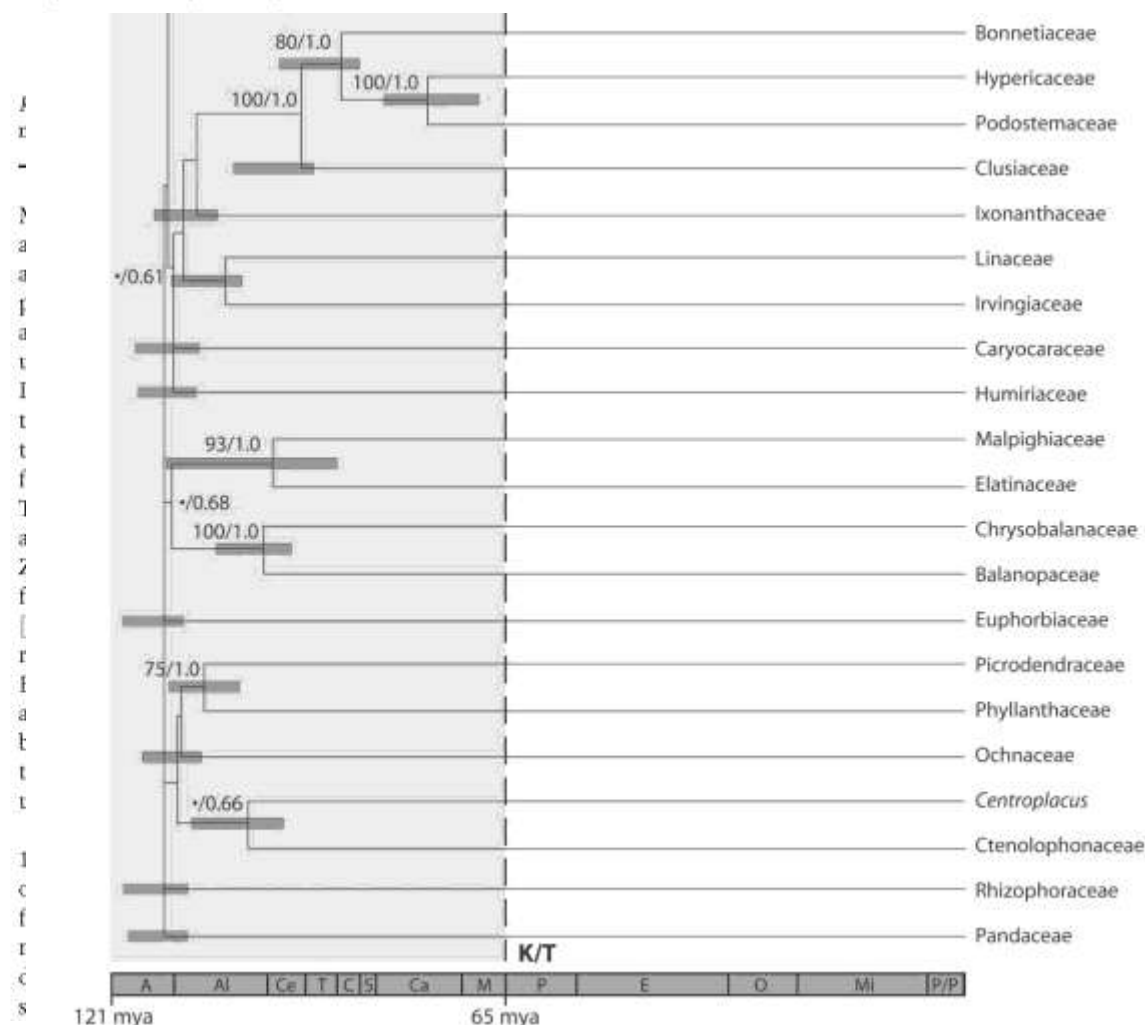
3. Department of Botany and Laboratories of Analytical Biology, Smithsonian Institution, P.O. Box 37012, National Museum of Natural History, MRC-166, Washington DC 20013-7012;

4. Biostratigraphy Team, Instituto Colombiano del Petróleo, AA 4185, Bucaramanga, Colombia

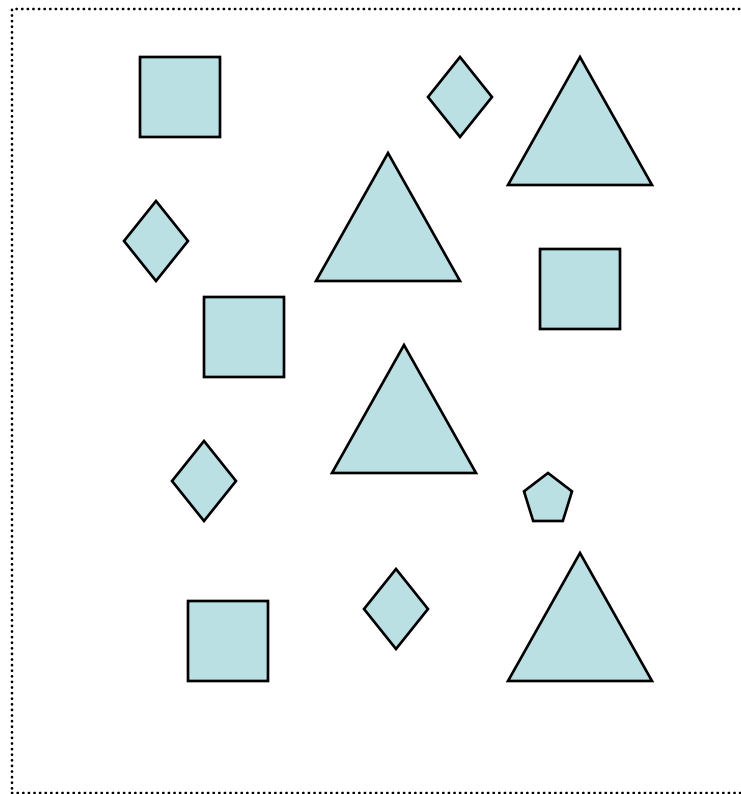
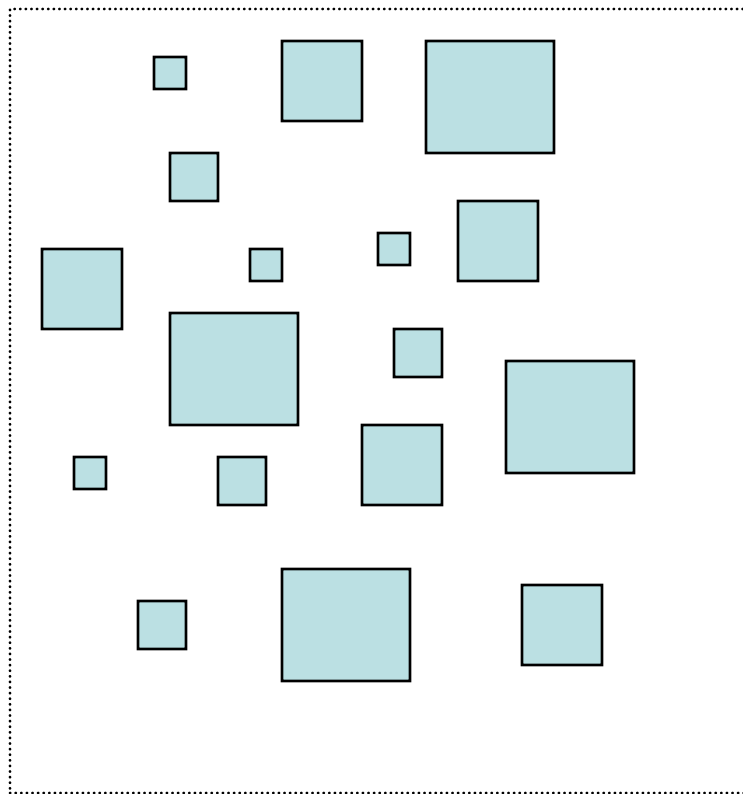
Submitted May 12, 2004; Accepted October 27, 2004;

Electronically published February 1, 2005

**ABSTRACT:** Fossil data have been interpreted as indicating that Late Cretaceous tropical forests were open and dry adapted and that modern closed-canopy rain forest did not originate until after the Cretaceous-Tertiary (K/T) boundary. However, some mid-Cretaceous leaf floras have been interpreted as rain forest. Molecular divergence-time estimates within the clade Malpighiales, which constitute a large percentage of species in the shaded, shrub, and small tree layer in tropical rain forests worldwide, provide new tests of these hypotheses. We estimate that all 28 major lineages (i.e., traditionally recognized families) within this clade originated in tropical rain forest well before the Tertiary, mostly during the Albian and Cenomanian (112–94 Ma). Their rapid rise in the mid-Cretaceous may have resulted from the origin of adaptations to survive and reproduce under a closed forest canopy. This pattern may also be paralleled by other similarly diverse lineages and supports fossil indications that closed-canopy tropical rain forests existed well before the K/T boundary. This case illustrates that dated phylogenies can provide an important new source of evidence bearing on the timing of major environmental changes, which may be especially useful when fossil evidence is limited or controversial.

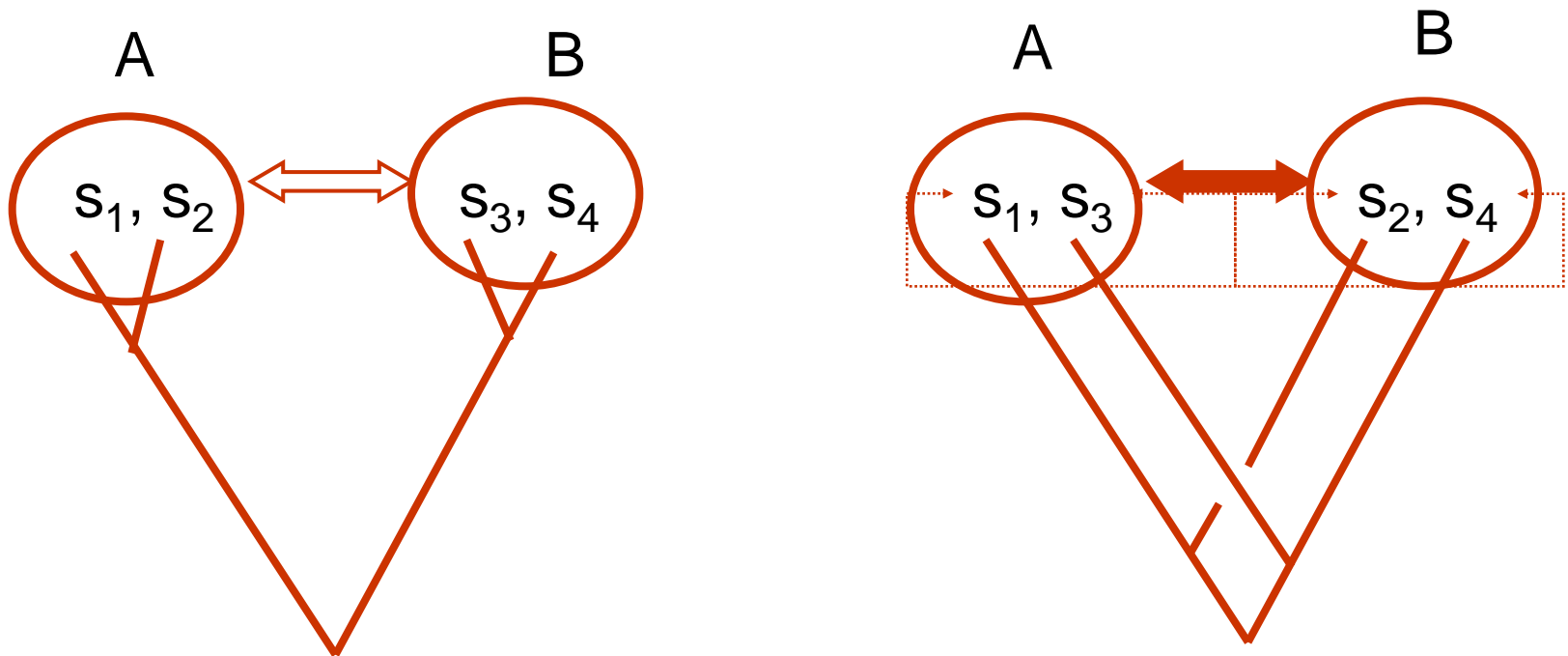


Going beyond species numbers



# Phylogenetic component on biodiversity

Measures of **phylogenetic diversity**



# Phylogenetic diversity in the Cape flora

Vol 445 | 15 February 2007 | doi:10.1038/nature05587

nature

## LETTERS

### Preserving the evolutionary potential of floras in biodiversity hotspots

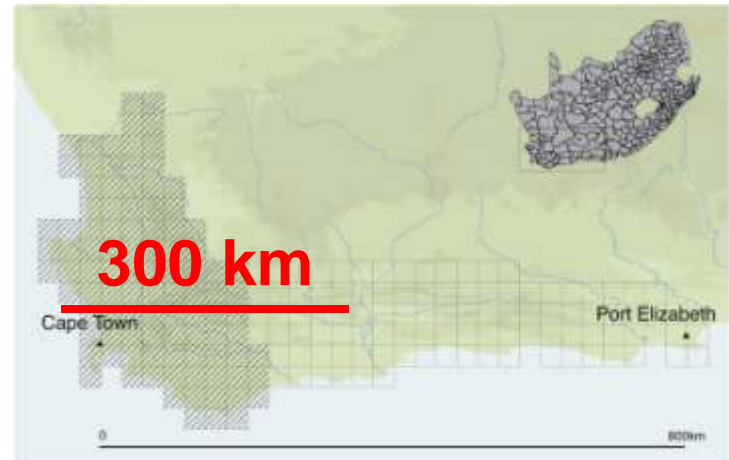
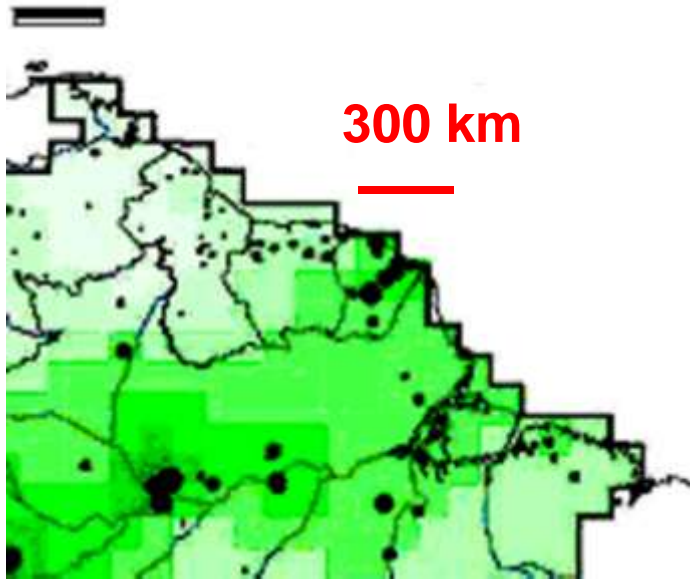
Félix Forest<sup>1,2,3\*</sup>, Richard Grenyer<sup>3\*</sup>, Mathieu Rouget<sup>4</sup>, T. Jonathan Davies<sup>5,6</sup>, Richard M. Cowling<sup>7</sup>, Daniel P. Faith<sup>8</sup>, Andrew Balmford<sup>9</sup>, John C. Manning<sup>1</sup>, Şerban Procheş<sup>10</sup>, Michelle van der Bank<sup>11</sup>, Gail Reeves<sup>1</sup>, Terry A. J. Hedderson<sup>2</sup> & Vincent Savolainen<sup>3</sup>

One of the biggest challenges for conservation biology is to provide conservation planners with ways to prioritize effort. Much attention has been focused on biodiversity hotspots<sup>1</sup>. However, the conservation of evolutionary process is now also acknowledged as a priority in the face of global change<sup>2</sup>. Phylogenetic diversity (PD) is a biodiversity index that measures the length of evolutionary pathways that connect a given set of taxa<sup>3,4</sup>. PD therefore identifies sets of taxa that maximize the accumulation of 'feature diversity'. Recent studies, however, concluded that taxon richness is a good surrogate for PD<sup>5-9</sup>. Here we show taxon richness to be decoupled from PD, using a biome-wide phylogenetic analysis of the flora of an undisputed biodiversity hotspot—the Cape of South Africa. We demonstrate that this decoupling has real-world importance for conservation planning. Finally, using a database of medicinal and economic plant use<sup>10</sup>, we demonstrate that PD protection is the best strategy for preserving feature diversity in the Cape. We should be able to use PD to identify those key regions that maximize future options, both for the continuing evolution of life on Earth and for the benefit of society.





# Scales



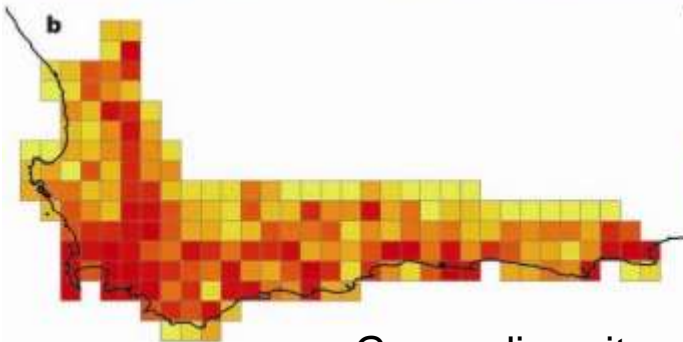
# Generating a phylogenetic hypothesis



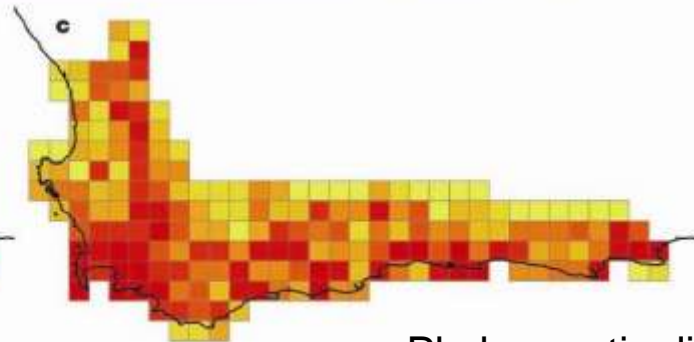
## Methods

**Sampling and DNA sequencing.** We sampled one exemplar species for 735 of the 943 genera of angiosperms currently recognized in the Cape (ca. 78%) and obtained sequence data for the plastid *rbcL* exon (ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit). Voucher information and GenBank/EMBL accession numbers are provided (Supplementary Information S2). Of the 735 *rbcL* sequences included here, 395 sequences were produced for this study and 340 were downloaded from GenBank/EMBL sequence databases (211 of which are represented by species present in the Cape). DNA was isolated using a modified version of the 2× CTAB method<sup>1</sup> and subsequently purified on cesium chloride/ethidium bromide gradients (1.55 g/ml density). PCR amplification of the *rbcL* exon was performed using primer combinations from Olmstead et al.<sup>2</sup>. PCR reactions (50 µl) were made with the

# Results



Genus diversity



Phylogenetic diversity

# Using phylogenies in ecology

1. Phylogenies provide a historical framework to ecological problems
2. Tools to obtain DNA sequences are readily available, easy to use, and they are cheap (10 euros/sequence) !!
3. Increasingly easy to develop tests of ecological theories that take into account phylogenetic information, using freely available statistical tools
4. For phylogeny reconstruction or manipulation: MrBayes and many others, Mesquite, the 'ape' package of the R software
5. Phylogenetic reconstruction is a different goal from barcoding methods (phylogenetics use highly conserved, long, sequences, while barcoding use much shorter, more variable sequences)

# Barcodes...



Short DNA sequences  $\approx 500\text{pb}$



Unique to a species

## Requirements:

Sufficiently conserved



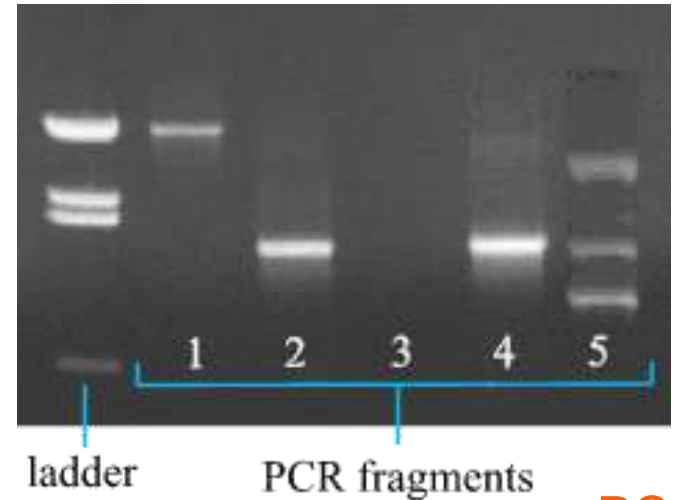
Universal primers

Easy alignment

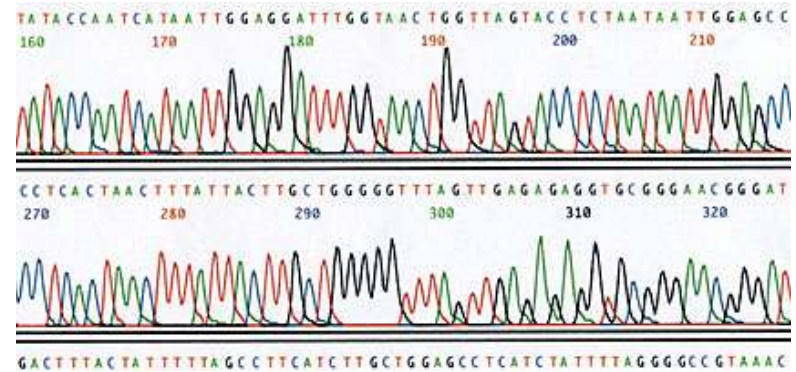
Enough variability



To enable discrimination  
between closely related  
species



PCR



Sequencing (two closely allied moth species) From Hebert, U Guelph.

# Work-package 3

- Sample at least one species of all of the 380 recognized tree genera in French Guiana, for sequences that are not already available in Genbank (ca. 200 for rbcL, 320 for matK).
- It will be best to have fresh material for these analyses, but if identifications are particularly problematic, herbarium specimen will be used
- One post-doctoral associate will be hired (starting in May) to begin this project. (S)he will be based in Toulouse, but will spend time in French Guiana (1-2 months) and in Kew (2 months).
- Expected outcome: a phylogeny resolved at the genus level for all the trees in French Guiana.