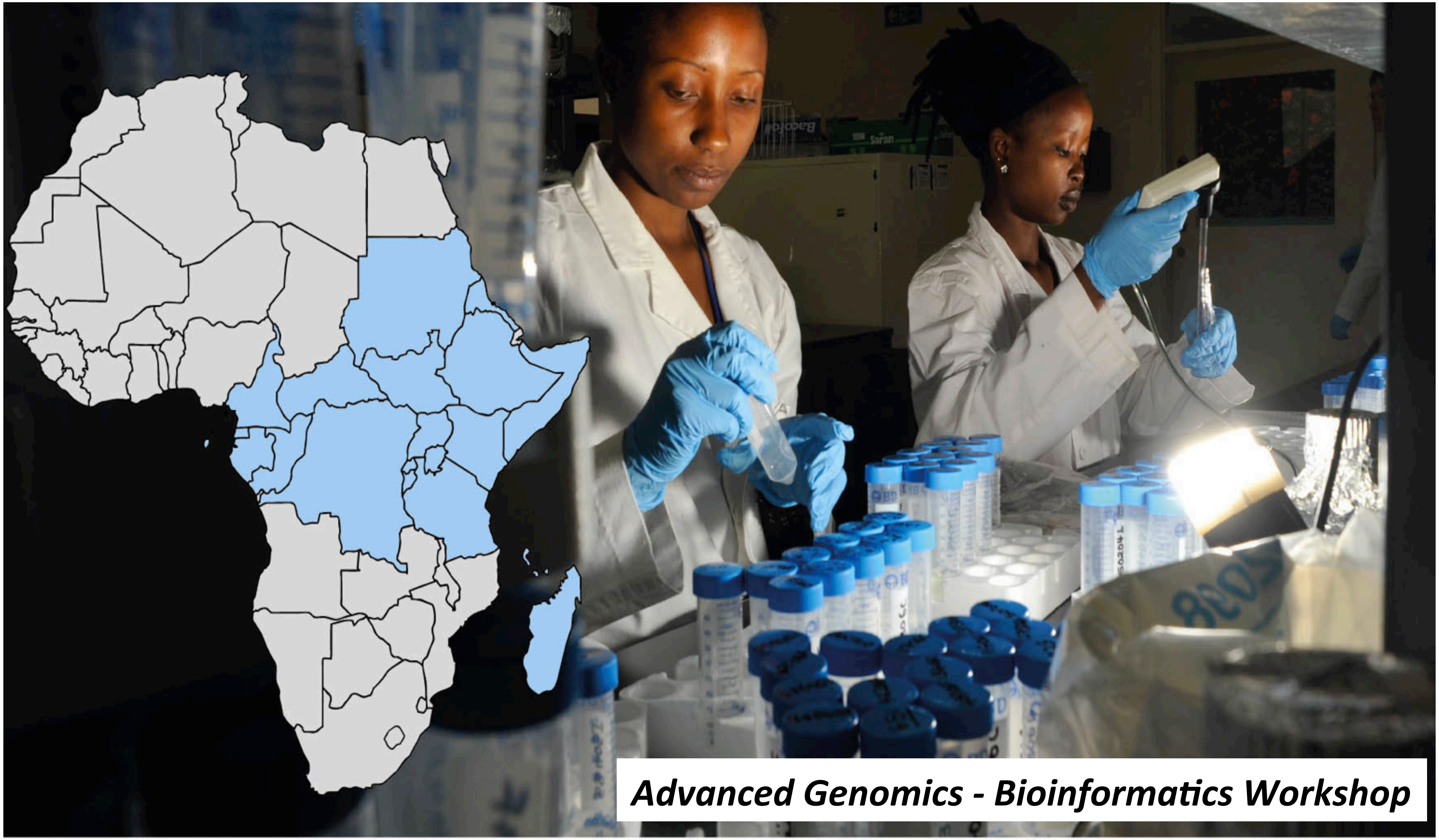


Advanced Genomics - Bioinformatics Workshop

Mark Wamalwa
BecA-ILRI Hub, Nairobi, Kenya
http://hub.africabiosciences.org/
m.wamalwa@cgiar.org

7th – 18th September 2015

ILRI INTERNATIONAL LIVESTOCK RESEARCH INSTITUTE

African Union

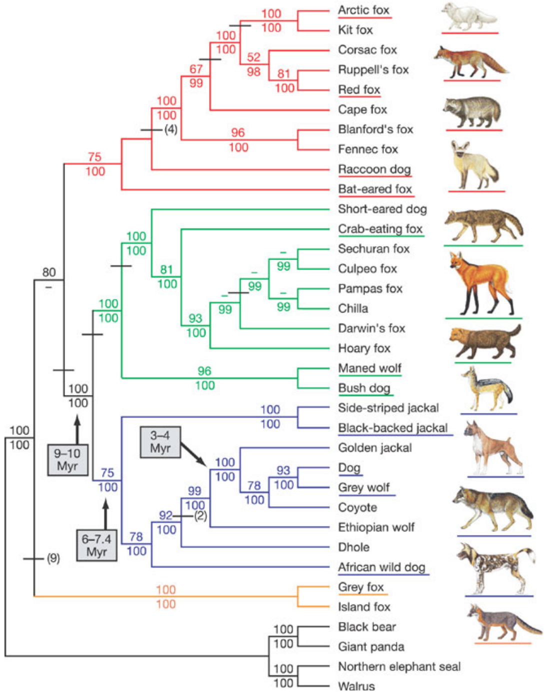
THE NEW PARTNERSHIP FOR AFRICA'S DEVELOPMENT NEPAD A PROGRAMME OF THE AFRICAN UNION

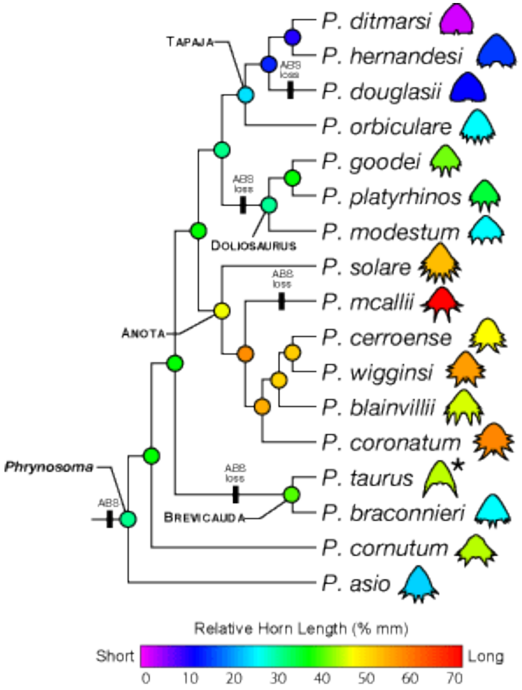
biosciences eastern and central africa

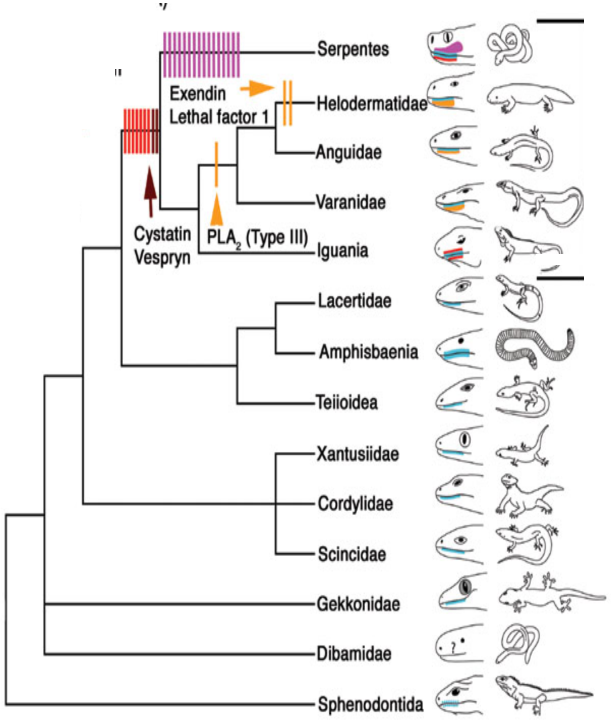
Phylogeny

- **Study of evolutionary relatedness among groups of organisms, achieved by:**
  - **Comparison of molecular data**
  - **Comparison of morphological data**

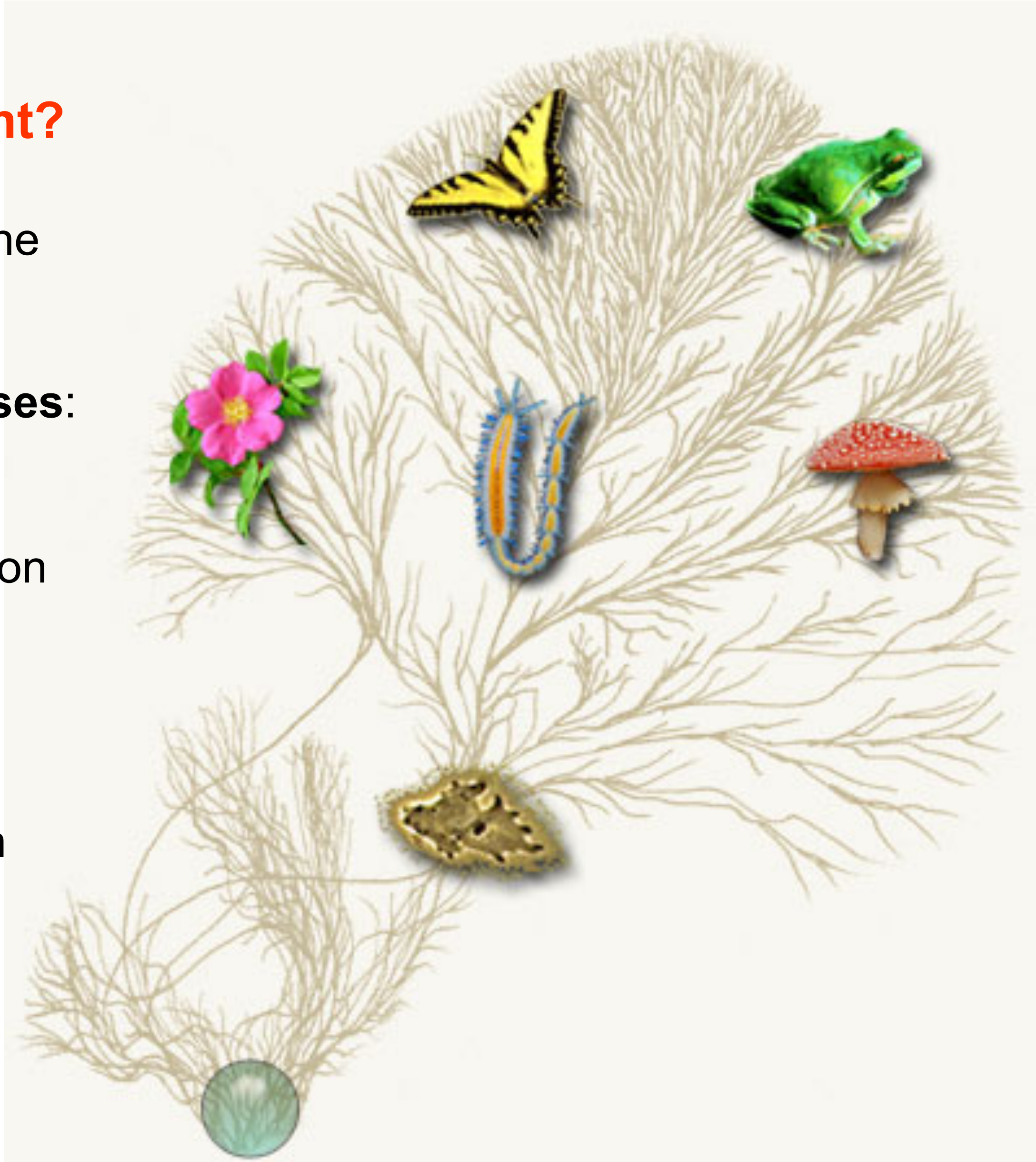
```
Outgroup    AAGCTTCATAGGAGCAACCATTCTAATAATAAGCCTCATAAAGCC
Species A   AAGCTTCACCGGCGCAGTTATCCTCATAATATGCCTCATAATGCC
Species B   GTGCTTCACCGACGCAGTTGTCCTCATAATGTGCCTCACTATGCC
Species C   GTGCTTCACCGACGCAGTTGCCCTCATGATGAGCCTCACTATGCA
```







Tree of life



Why is phylogeny important?

Understanding and classifying the diversity of life on Earth

Testing evolutionary hypotheses:

- trait evolution
- coevolution
- mode and pattern of speciation
- correlated trait evolution
- biogeography
- geographic origins
- age of different taxa
- nature of molecular evolution
- disease epidemiology

…and many more applications!

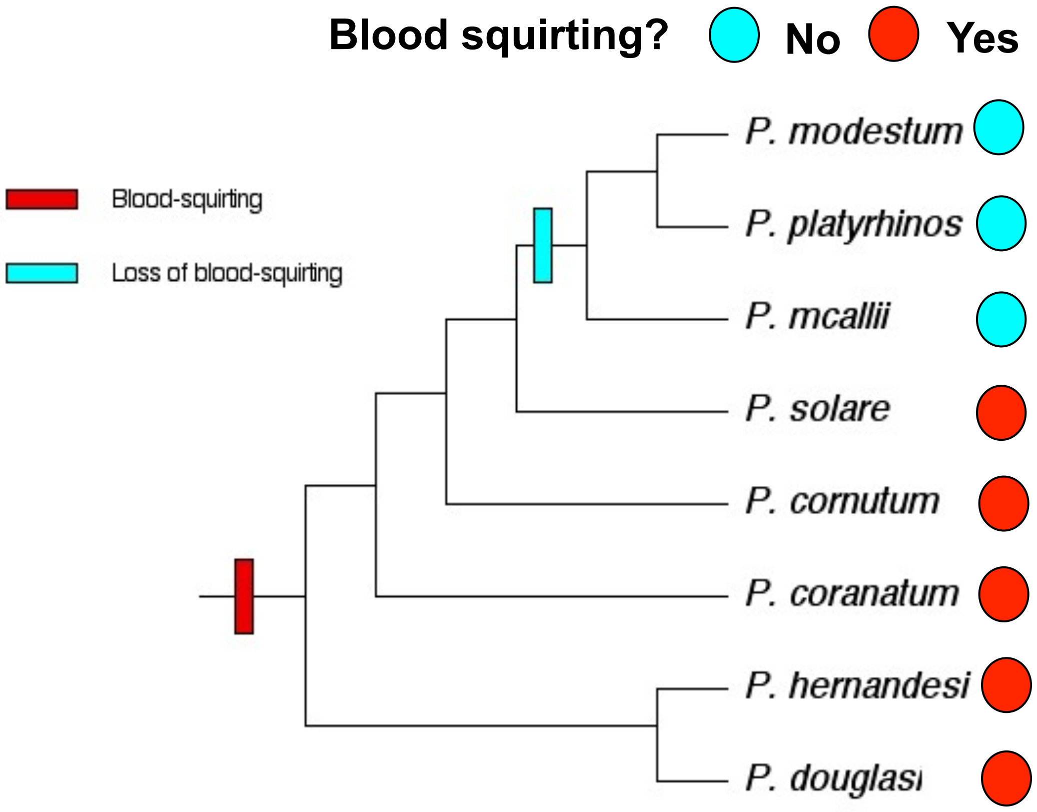
Testing evolutionary hypotheses

Mapping evolutionary transitions

Some horned lizards squirt blood from their eyes when attacked by canids

How many times has blood-squirting evolved?

This phylogeny suggests a single evolutioary gain and a single loss of blood squirting



Blood squirting? 🔵 **No** 🔴 **Yes**

🟥 Blood-squirting

🟦 Loss of blood-squirting

P. modestum 🔵

P. platyrhinos 🔵

P. mcallii 🔵

P. solare 🔴

P. cornutum 🔴

P. coranatum 🔴

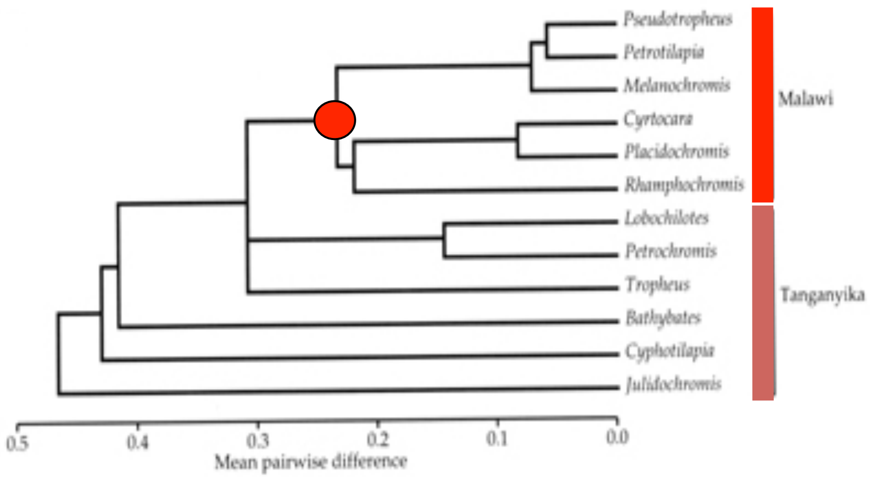
P. hernandesi 🔴

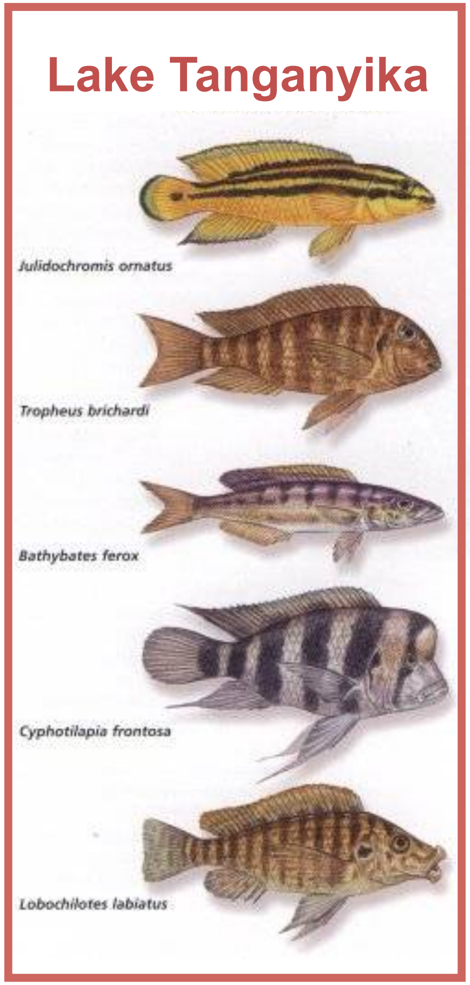
P. douglasi 🔴

Testing evolutionary hypotheses

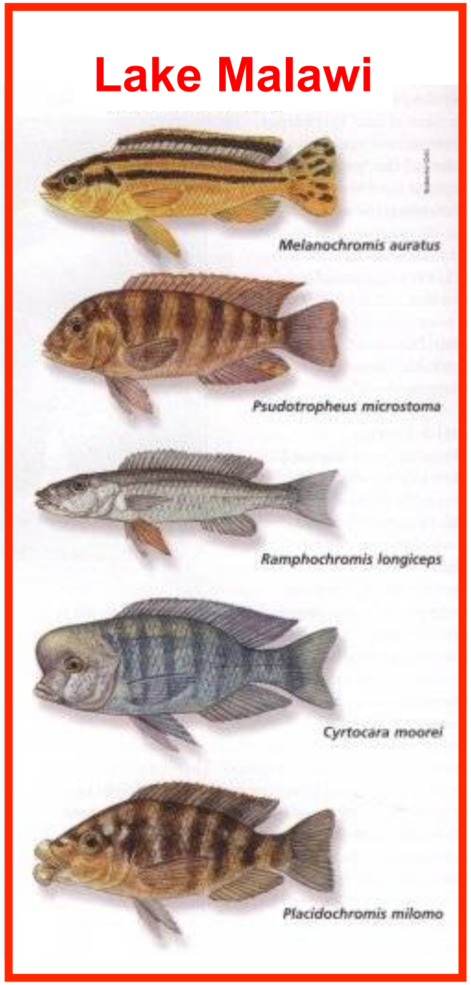
Convergence and modes of speciation

What can this phylogeny tell us about homology/analogy and speciation?





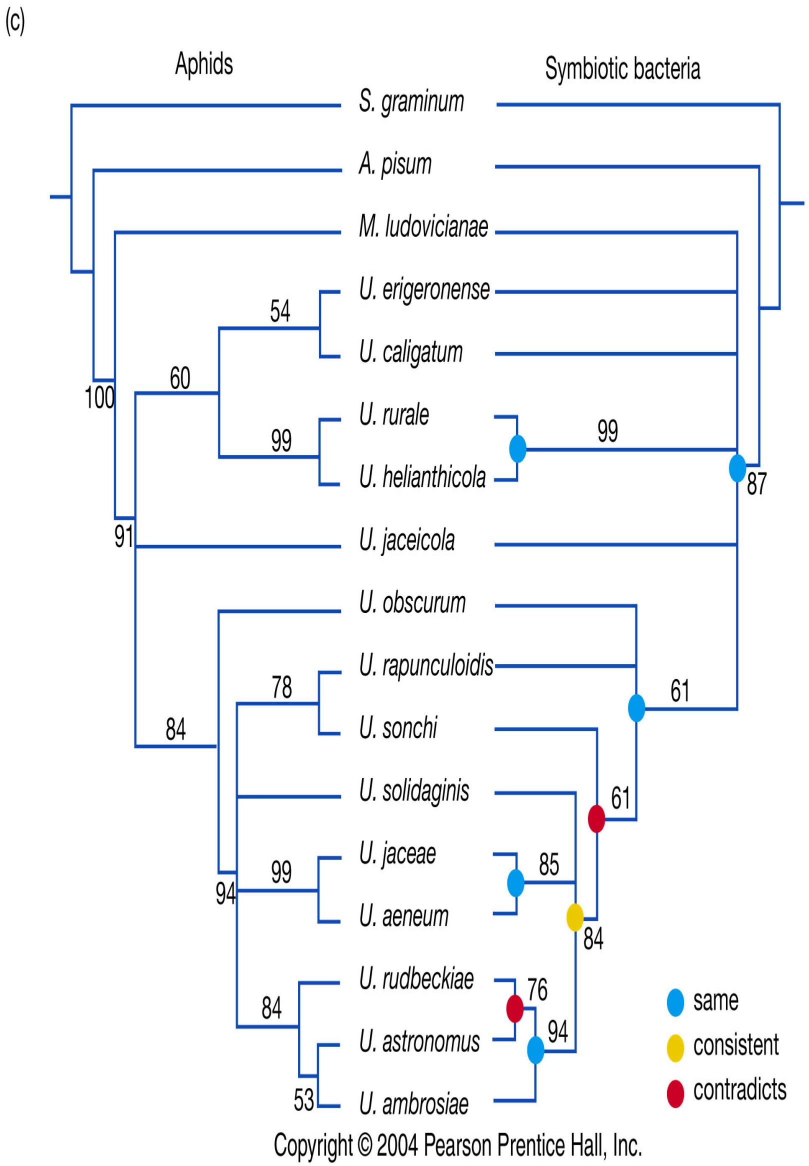
Lake Tanganyika



Lake Malawi

1. Similarities between each pair are the result of **convergence**
2. **Sympatric speciation** more likely than allopatric speciation

Testing evolutionary hypotheses



Copyright © 2004 Pearson Prentice Hall, Inc.

Clark et al. (2000)

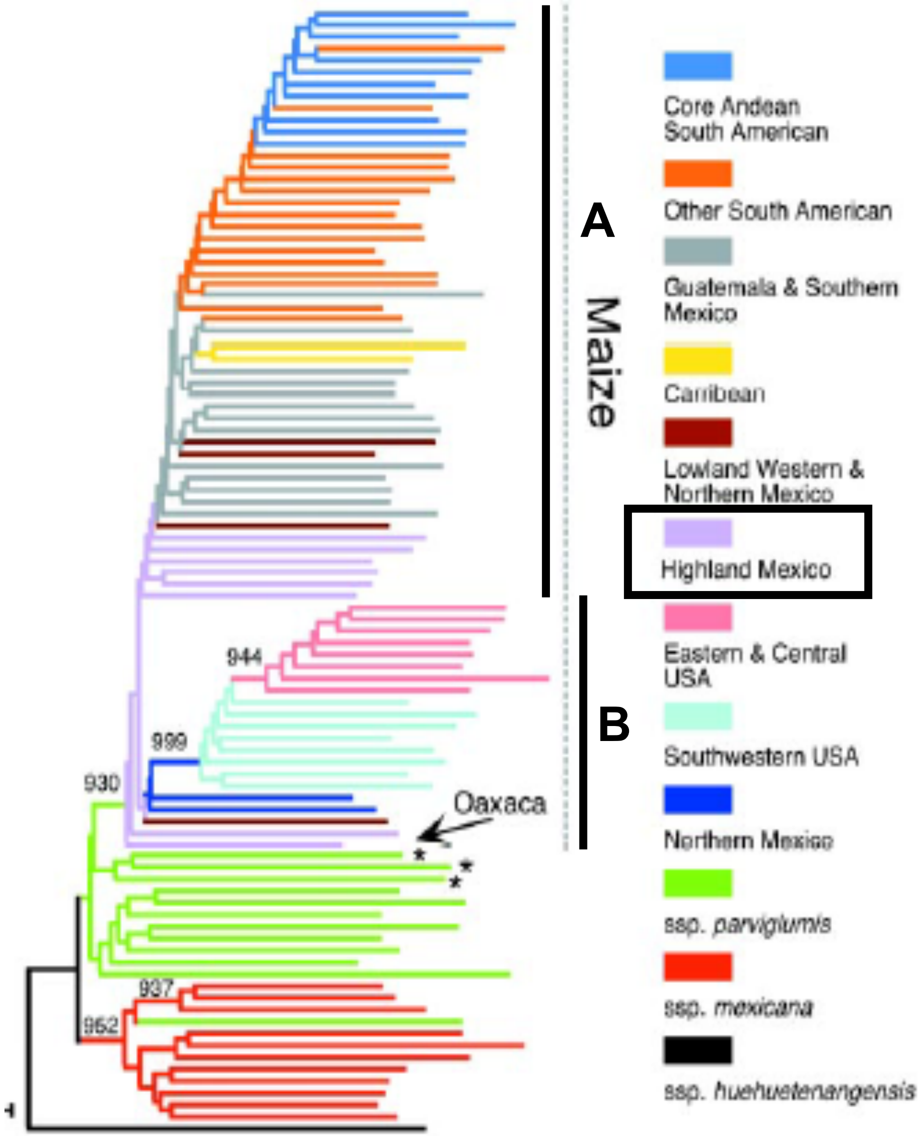
Coevolution

Aphids and bacteria are symbiotic

Given this close relationship, we might expect that speciation in an aphid would cause parallel speciation in the bacteria

When comparing phylogenies for each group we see evidence for **reciprocal cladogenesis** (but also contradictions)

Testing evolutionary hypotheses



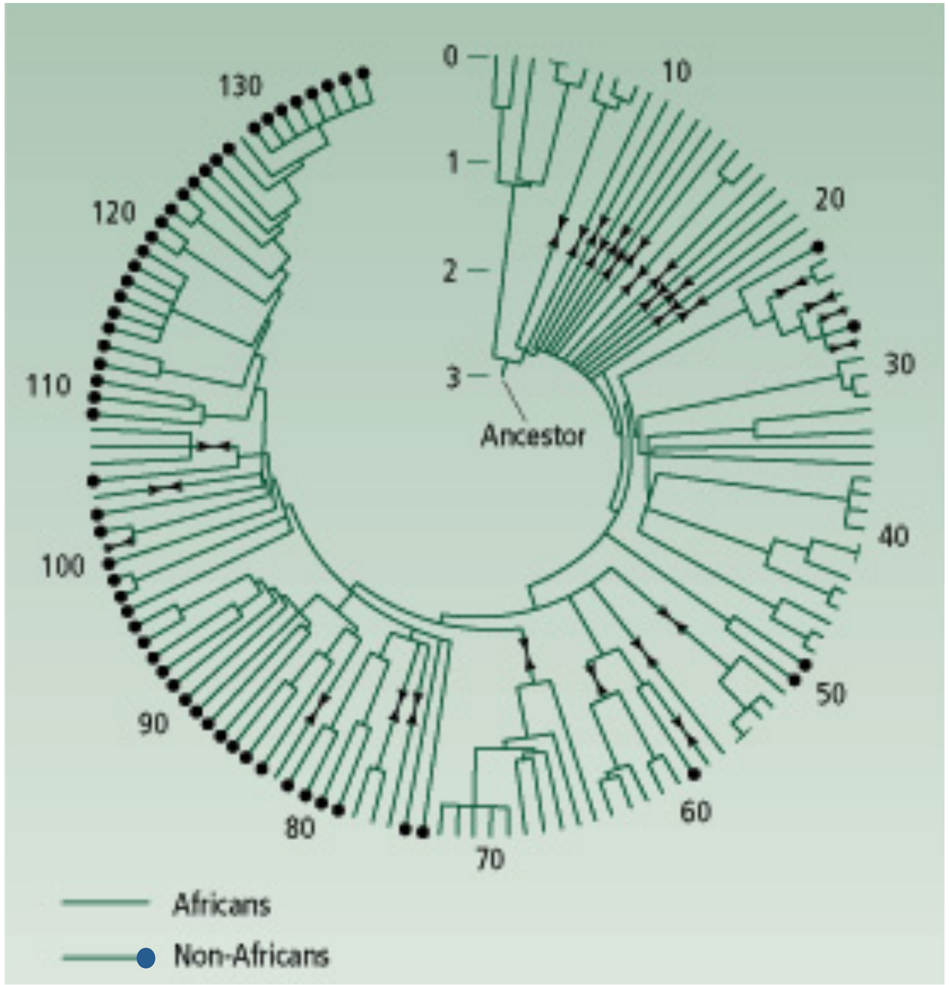
Matsuoka et al. (2002)

Geographic origins

Where did domestic corn (*Zea mays maize*) originate?

Populations from Highland Mexico are at the base of each maize clade

Testing evolutionary hypotheses



Vigilant et al. (1991) *Science*

Geographic origins

Where did humans originate?

Each tip is one of 135 different mitochondrial DNA types found among 189 individual humans

African mtDNA types are clearly basal on the tree, with the non-African types derived

Suggests that humans originated in Africa

What Sequences to Study?

- Study more than just one region!

- Different sequences evolve at different rates

  - Proteins
  - Highly variable sequences (ex: immunoglobulin genes)
  - Highly conserved (actin, rRNA coding regions)
  - Different regions within a single gene can evolve at different rates (conserved vs. variable domains)

Molecular Phylogenies

- The gene compared must evolve at a rate comparable to the divergence time of the organism; for example:
  - 18S rRNA gene for phylum-level divergences since it evolves very slowly.
  - Hemoglobin genes for mammalian orders.
  - Mitochondrial DNA for species divergences within a genus.
  - Repetitive DNA sequences (e.g. microsatellites) for individuals within species.

Tree Building Goals

- Maximize shared derived character states in a lineage
- Minimize homoplasies
  - Parallel changes, convergences, and reversals of character states between and within lineages

Tree building methods

- Distance methods
- Maximum Parsimony
- Maximum likelihood
- Bayesian inference

All methods can be used with different substitution models

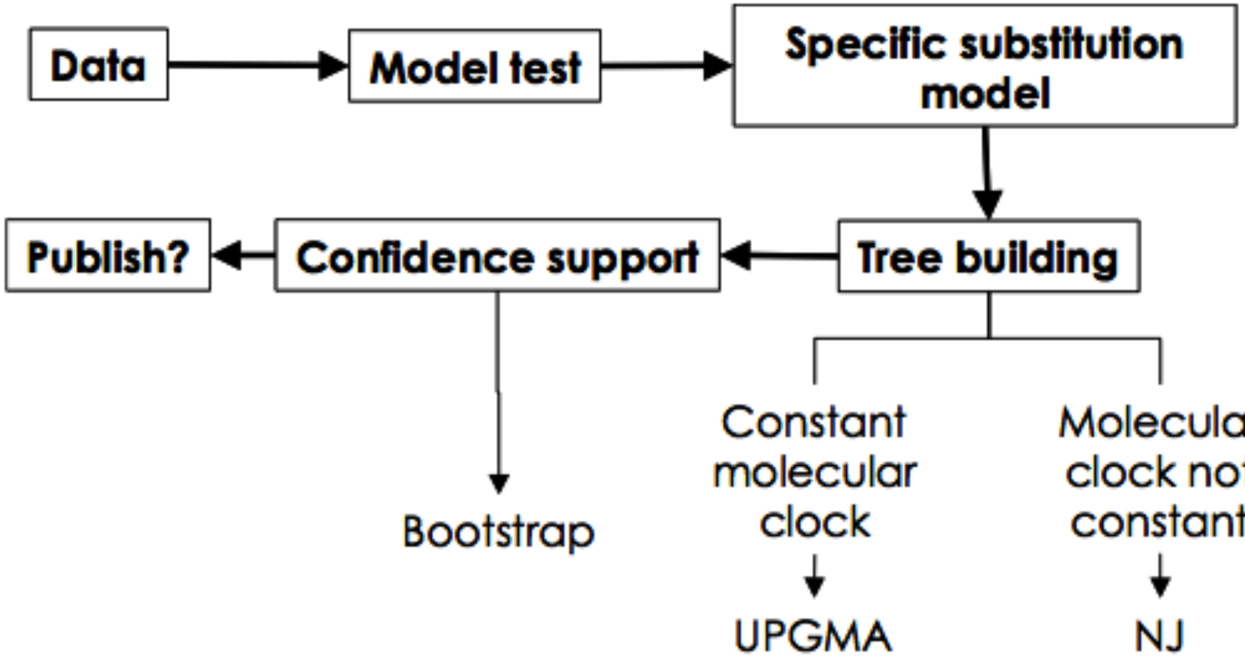
Distance methods

- UPGMA (Unweighted Pair Group Method with Arithmetic mean): same rate of evolution on each branch

- The Neighbor Joining method = most popular method

  does not assume the same rate of evolution on each branch of a tree

Distance methods: Procedure



Building Trees with Parsimony

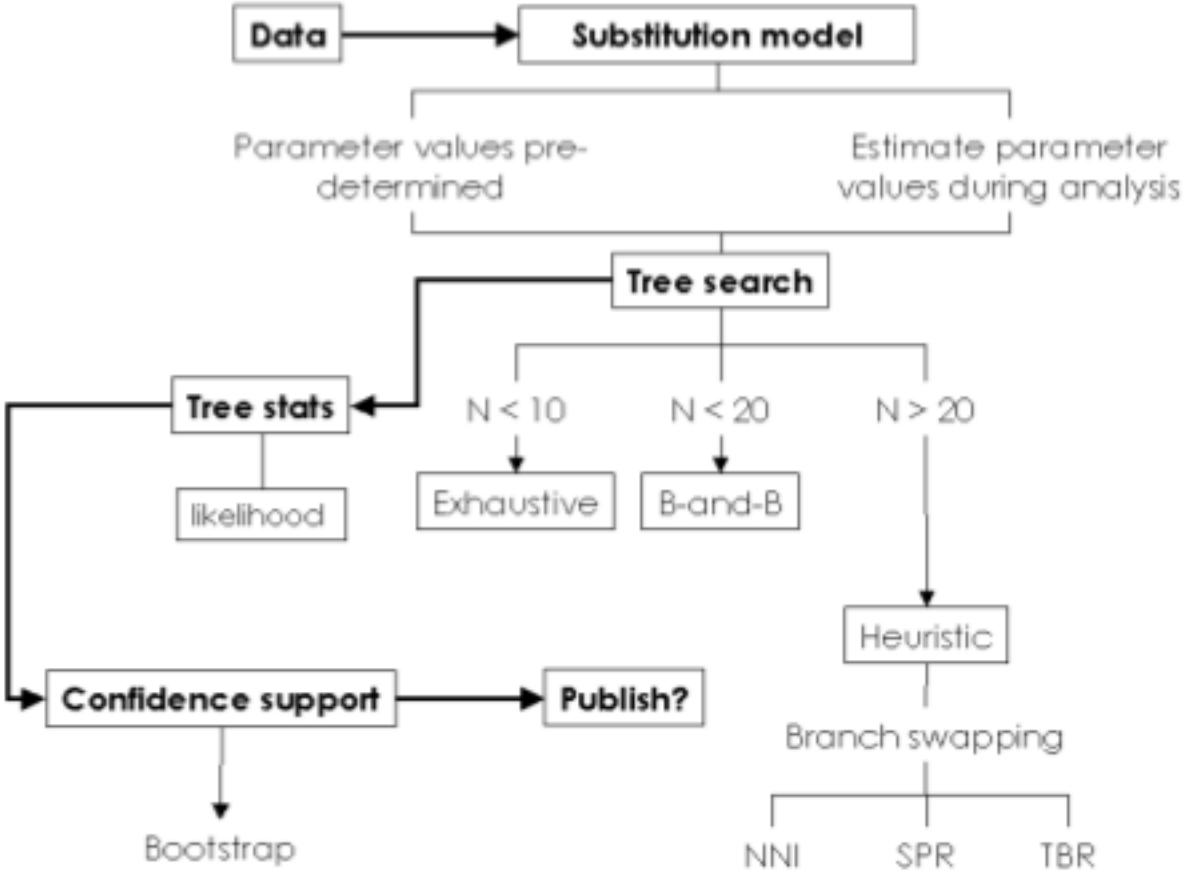
- **Parsimony** involves evaluating all possible trees and giving each a score based on the number of evolutionary changes that are needed to explain the observed data.

- The best tree is the one that requires the fewest base changes for all sequences to derive from a common ancestor.

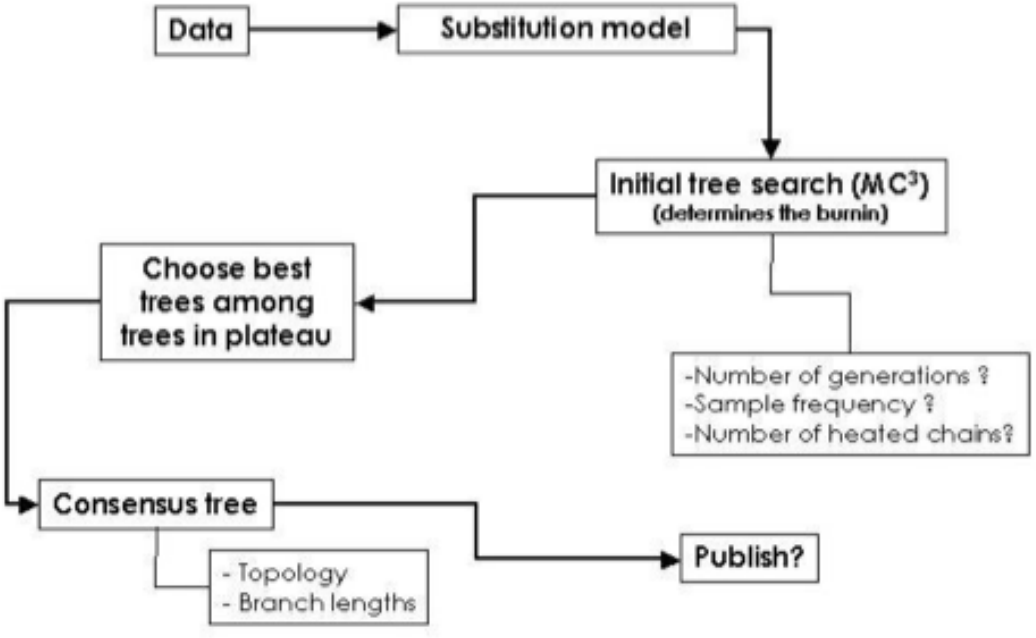
Maximum likelihood and bayesian methods

- Allows for substitution rates to differ on lineages and sites: appropriate for distantly related species
- Estimates the likelihood of a tree = probability of the data given an evolutionary model
- Complex and computationally intensive!

Maximum likelihood: Procedure



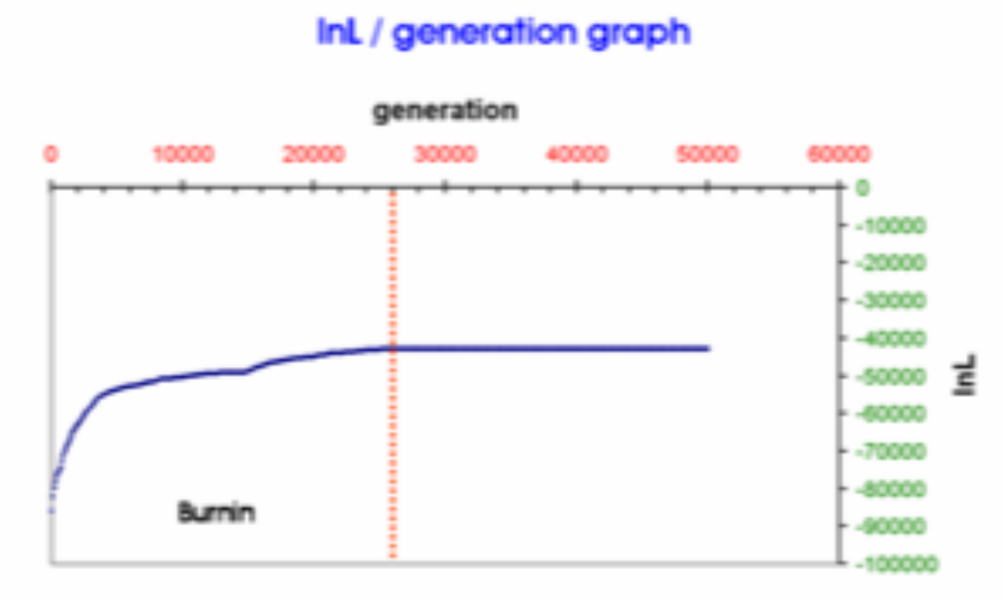
Bayesian methods: Procedure



Bayes theorem in phylogenetic terms

$$P[Tree \mid Data] = \frac{P[Data \mid Tree] x P[Tree]}{P[Data]}$$

Bayesian methods: BURN-IN



- Burn-in refers to the exclusion of trees before stationary state is reached.
- Likelihood values converge after 26000 generations. A tree was saved every 100th generation.
- The number of trees to be exluded = 2600 (i.e. burn-in value).

USING MRBAYES

- MrBayes is developed by Huelsenbeck and Ronquist (2001).
- It is used to determine the topology of trees and to estimate the posterior probability of both the topologies and clades.
- Infile format:

✓ Datafile in the NEXUS format

✓ MrBayes block included below the data matrix

✓ Nexus file in the same directory as the MrBayes executable file

Many Thanks