## Algorithmic issues in (co)phylogenetic analysis

blerina sinaimeri

evolution

This is NOT Your Family Tree Great


This is Your Family Tree


This is NOT Evolution


This is Evolution
Mammals


Cattoon by Mathew Botavn of Mucomb, II, vith Kind pemsission of Flocida Citisend for Scienct, Sejt. 2010

## Phylogenetic tree



Phylogenetic Tree

## Phylogenetic tree



Phylogenetic Tree

## Phylogenetic Trees



- rooted / unrooted
- binary / k-ary
- labeled from a set / labeled from a multiset
- unweighted / weighted (branch lengths)
- unordered / ordered



## Phylogenetic Trees

- Maximum Parsimony
- Maximum Likelihood method


## Maximum Parsimony

Sequence 1 TGC<br>Sequence 2 TAC<br>Sequence 3 AGG<br>Sequence 4 AAG



Find the "best" tree...but what does "best" mean?

## Maximum Parsimony

Sequence 1 TGC
Sequence 2 TAC
Sequence 3 AGG
Sequence 4 AAG

Find the "best" tree...but what does "best" mean?

In Maximum Parsimony: Minimize the number of mutations across the edges


## Maximum Parsimony

Sequence 1 TGC<br>Sequence 2 TAC<br>Sequence 3 AGG<br>Sequence 4 AAG



Find the "best" tree...but what does "best" mean?

In Maximum Parsimony: Minimize the number of mutations across the edges


## Maximum Parsimony



## Maximum Parsimony

## The problem

- Input : n DNA sequences
- Goal: Find the tree that minimizes the number of mutations along the edges.

Check for every tree?

$$
\text { Possible unrooted trees } \frac{(2 n-5)!}{2^{n-3}(n-3)!}
$$

## Maximum Parsimony

## The problem

- Input : n DNA sequences
- Goal: Find the tree that minimizes the number of mutations along the edges.

Finding one optimal tree is NP-hard!

## Maximum Likelihood

- Given certain rules about how sequences change over time, the best tree should reflect the most likely sequence of evolutionary events.
- maximize the probability that a given tree could have produced the observed data (i.e., the likelihood)


## Differences with the parsimonious method

- Use of an explicit evolutionary model
- Allows variable substitution rates for each branch


## Comparing Trees

How similar two phylogenetic trees are?

- Robinson-Foulds
- Triplet distance
- Maximum agreement subtree
- Edit distances (SPR, TBR, NNI)


## Tree Metrics

## SPR (Subtree Prune and Regraft)



The SPR distance (dsPR) is the minimal number of moves that transforms one tree into the other.

NP-hard
3-approximation algorithm

## Maximum Agreement Forest

Technical detail


## Maximum Agreement Forest

A Forest of T is a disjoint collection of phylogenetic subtrees whose union of leaf sets is Xup.



## Maximum Agreement Forest



## Maximum Agreement Forest



## Maximum Agreement Forest

$m\left(T_{1}, T_{2}\right)=$ size of maximum agreement forest

Theorem. (BS04)
Let $T_{1}$ and $T_{2}$ be two binary phylogenetic $X$-trees. Then

$$
\operatorname{dspR}_{\mathrm{sPR}}\left(\mathrm{~T}_{1}, \mathrm{~T}_{2}\right)=\mathrm{m}\left(\mathrm{~T}_{1}, \mathrm{~T}_{2}\right)-1
$$

## Exercise 1

- Explain the role of $\rho$ in the relation between SPR and MAF. Does the theorem hold without introducing $\rho$ ?


## Maximum Agreement Forest

3-approximation algorithm for MAF


## coevolution

## Symbiosis

## Mutualism



Human Microbiota

Parasitism


Gopher


Lice

## Interspecific interaction



Plant diversity


Mimicry


Parasitism


Human Microbiota


Mutualism

