### Algorithmic issues in (co)phylogenetic analysis

blerina sinaimeri

## evolution



Cartoon by Matthew Bonnan of Macomb, IL, with kind pennission of Florida Citizens for Science, Sept. 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

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

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



- 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



- 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



- 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



#### The problem

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

Check for every tree? Possible unrooted trees  $\frac{(2n-5)!}{2^{n-3} (n-3)!}$ 

#### 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 ( $d_{\text{SPR}}$ ) is the minimal number of moves that transforms one tree into the other.

NP-hard 3-approximation algorithm

#### **Technical detail**





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









 $m(T_1,T_2)$ = size of maximum agreement forest

Theorem. (BS04)

Let T<sub>1</sub> and T<sub>2</sub> be two binary phylogenetic X-trees. Then

 $d_{SPR}(T_1,T_2) = m(T_1,T_2)-1$ 

## Exercise 1

 Explain the role of ρ in the relation between SPR and MAF. Does the theorem hold without introducing ρ?

**3-approximation algorithm for MAF** 



## coevolution

## Symbiosis

#### Mutualism

#### Parasitism







Human Microbiota

Gopher

Lice

## Interspecific interaction



Plant diversity





Parasitism



#### Human Microbiota



Mutualism