

Lecture 1.3:

Phylogenetic Methods

Popular phylogenetic methods

1. Maximum parsimony
2. Distance-based methods
3. Maximum likelihood
4. Bayesian inference

Model-based methods



Distance-Based Methods

Distance-based methods

brown bear CGTTAGTACACT
cave bear CGATAGTTCACACT
black bear CGTTAGTTTACC
giant panda CATTGGTTTACT

MODEL



	brown bear	cave bear	black bear	giant panda
brown bear	-			
cave bear	.1	-		
black bear	.3	.3	-	
giant panda	.4	.5	.4	-

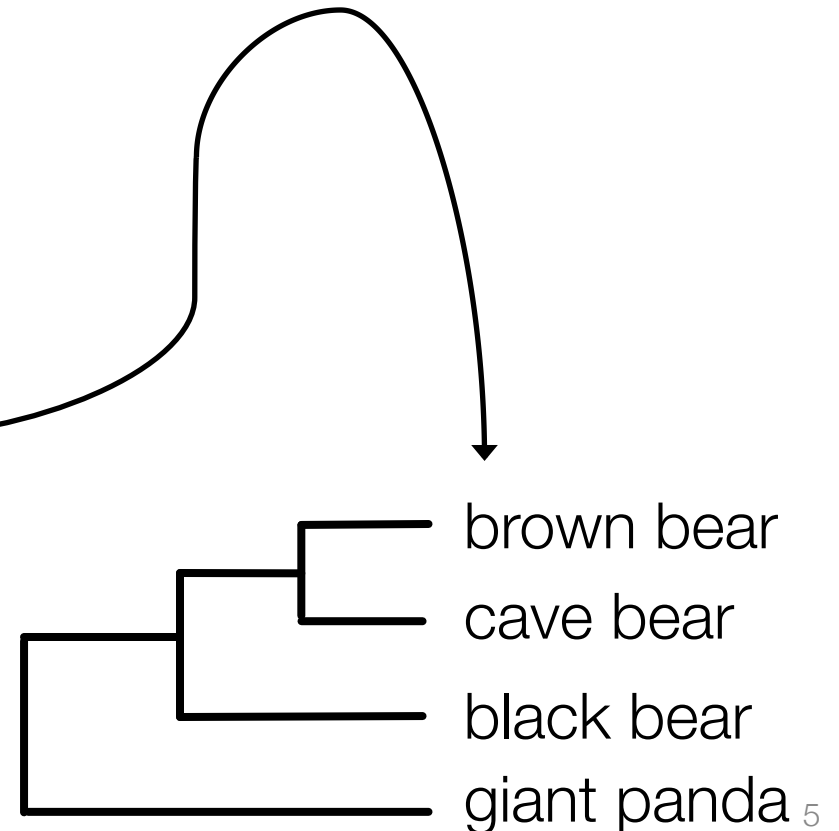
Neighbour joining

brown bear CGTTAGTACACT
cave bear CGATAGTTCACACT
black bear CGTTAGTTTACC
giant panda CATTGGTTTACT

MODEL

	brown bear	cave bear	black bear	giant panda
brown bear	-			
cave bear	.1	-		
black bear	.3	.3	-	
giant panda	.4	.5	.4	-

CLUSTERING
ALGORITHM



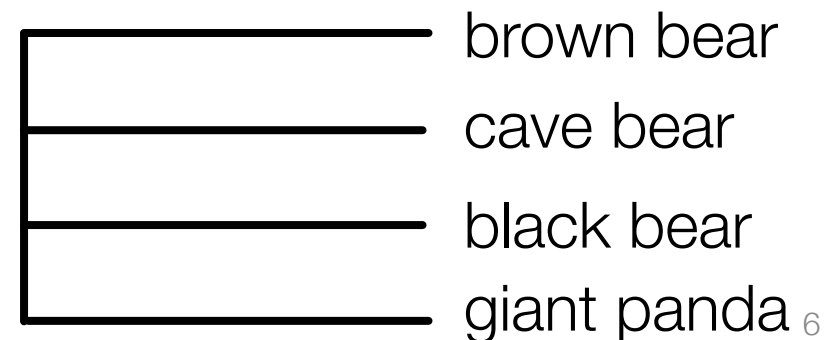
Neighbour joining

brown bear CGTTAGTACACT
cave bear CGATAGTTCACCT
black bear CGTTAGTTTACC
giant panda CATTGGTTTACT

MODEL

	brown bear	cave bear	black bear	giant panda
brown bear	-			
cave bear	.1	-		
black bear	.3	.3	-	
giant panda	.4	.5	.4	-

CLUSTERING
ALGORITHM



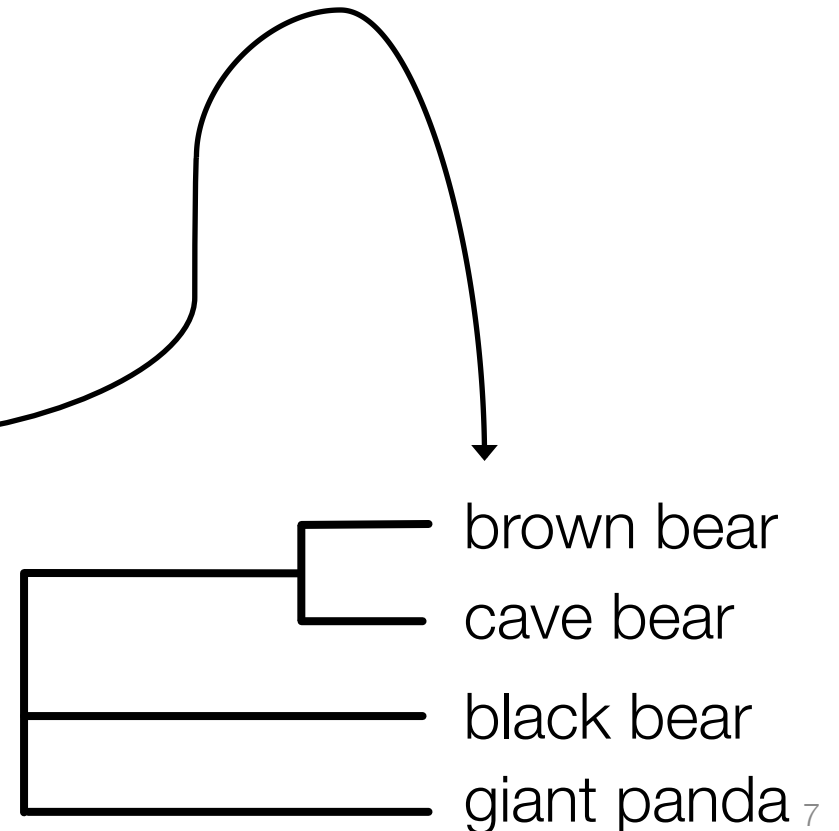
Neighbour joining

brown bear CGTTAGTACACT
cave bear CGATAGTTCACACT
black bear CGTTAGTTTACC
giant panda CATTGGTTTACT

MODEL

	brown bear	cave bear	black bear	giant panda
brown bear	-			
cave bear	.1	-		
black bear	.3	.3	-	
giant panda	.4	.5	.4	-

CLUSTERING
ALGORITHM



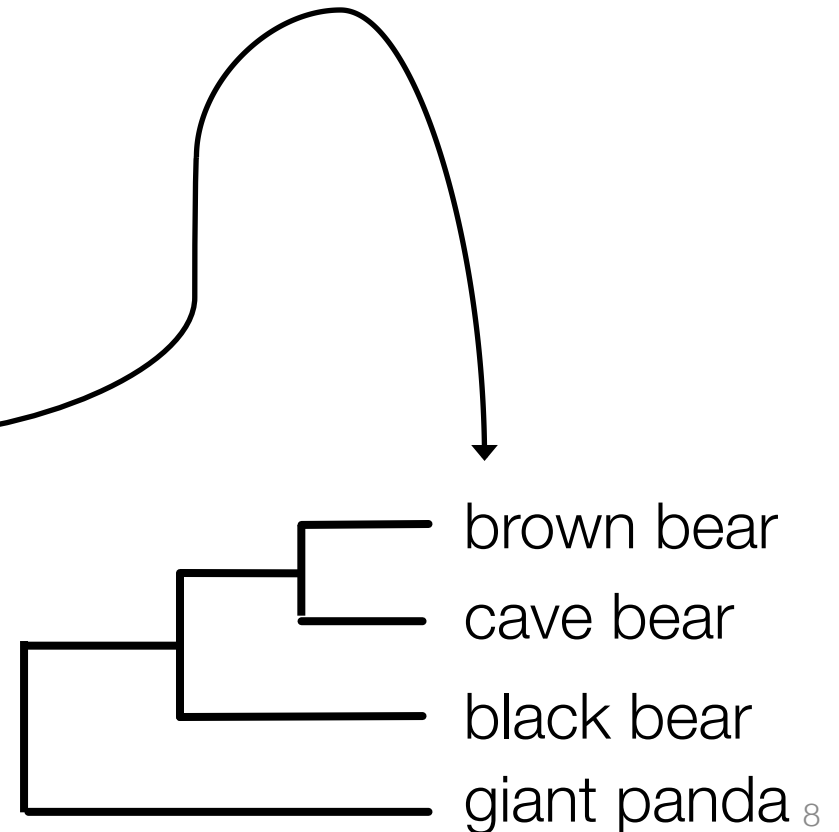
Neighbour joining

brown bear CGTTAGTACACT
cave bear CGATAGTTCACACT
black bear CGTTAGTTTACC
giant panda CATTGGTTTACT

MODEL

	brown bear	cave bear	black bear	giant panda
brown bear	-			
cave bear	.1	-		
black bear	.3	.3	-	
giant panda	.4	.5	.4	-

CLUSTERING
ALGORITHM



Distance-based methods

- Clustering algorithms
 - Unweighted Pair Group Method with Arithmetic Mean (UPGMA)
 - Neighbour joining
- Tree searching using optimality criteria
 - Minimum evolution
 - Least-squares inference

Strengths and weaknesses

- **Strengths**

- Very quick method
- Deals with multiple substitutions and long-branch attraction

- **Weaknesses**

- Does not use all information in alignment
- Loss of information in pairwise comparisons
- Unable to implement sophisticated evolutionary models

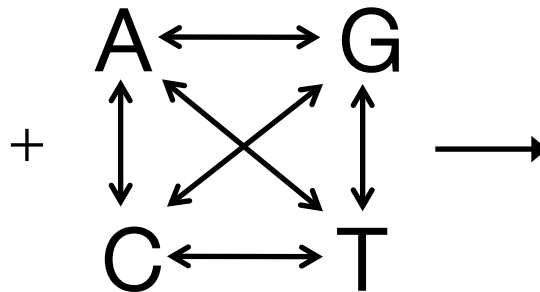
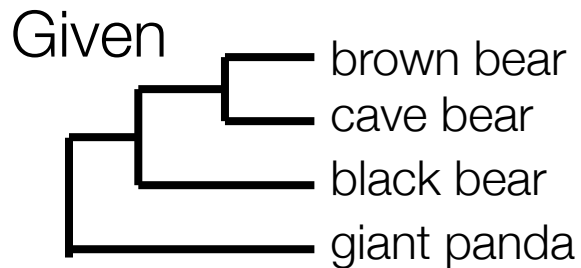
Maximum Likelihood

Maximum likelihood

Likelihood of hypothesis $H =$

$$P(D | H)$$

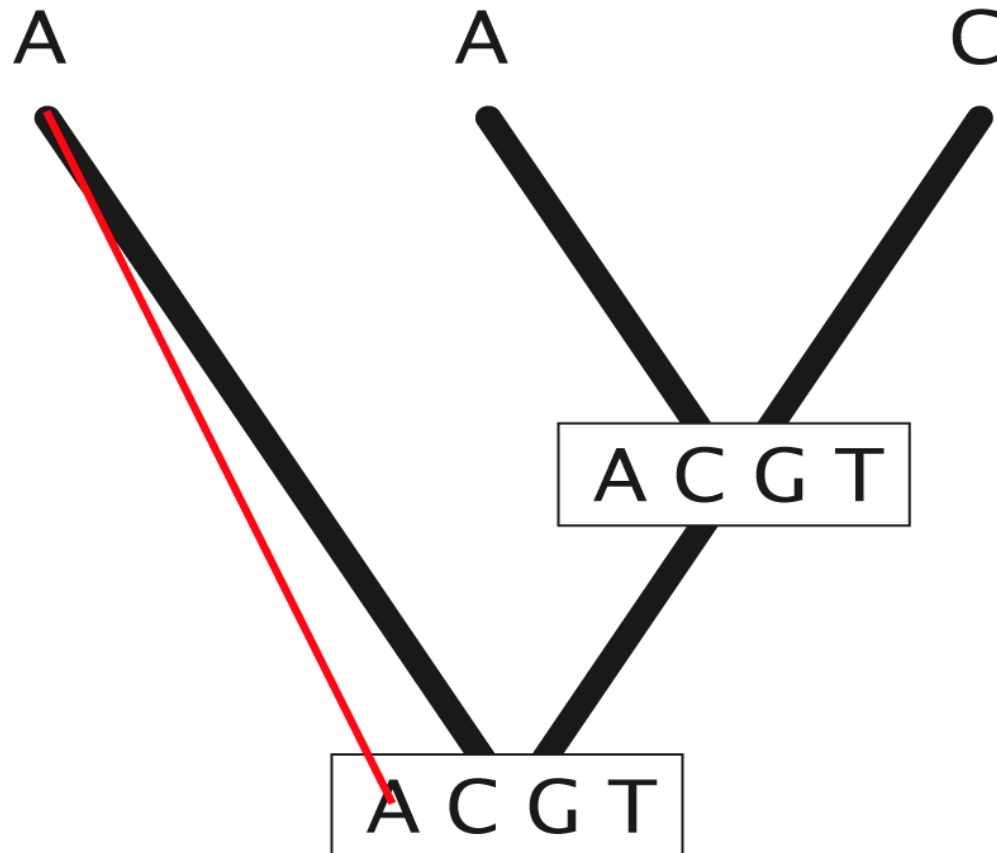
the probability of the data, given the hypothesis



Probability of?

Brown bear	CGTTAGTACACT
Cave bear	CGATAGTTCACACT
Black bear	CGTTAGTTTACC
Giant panda	CATTGGTTTACT

Maximum likelihood

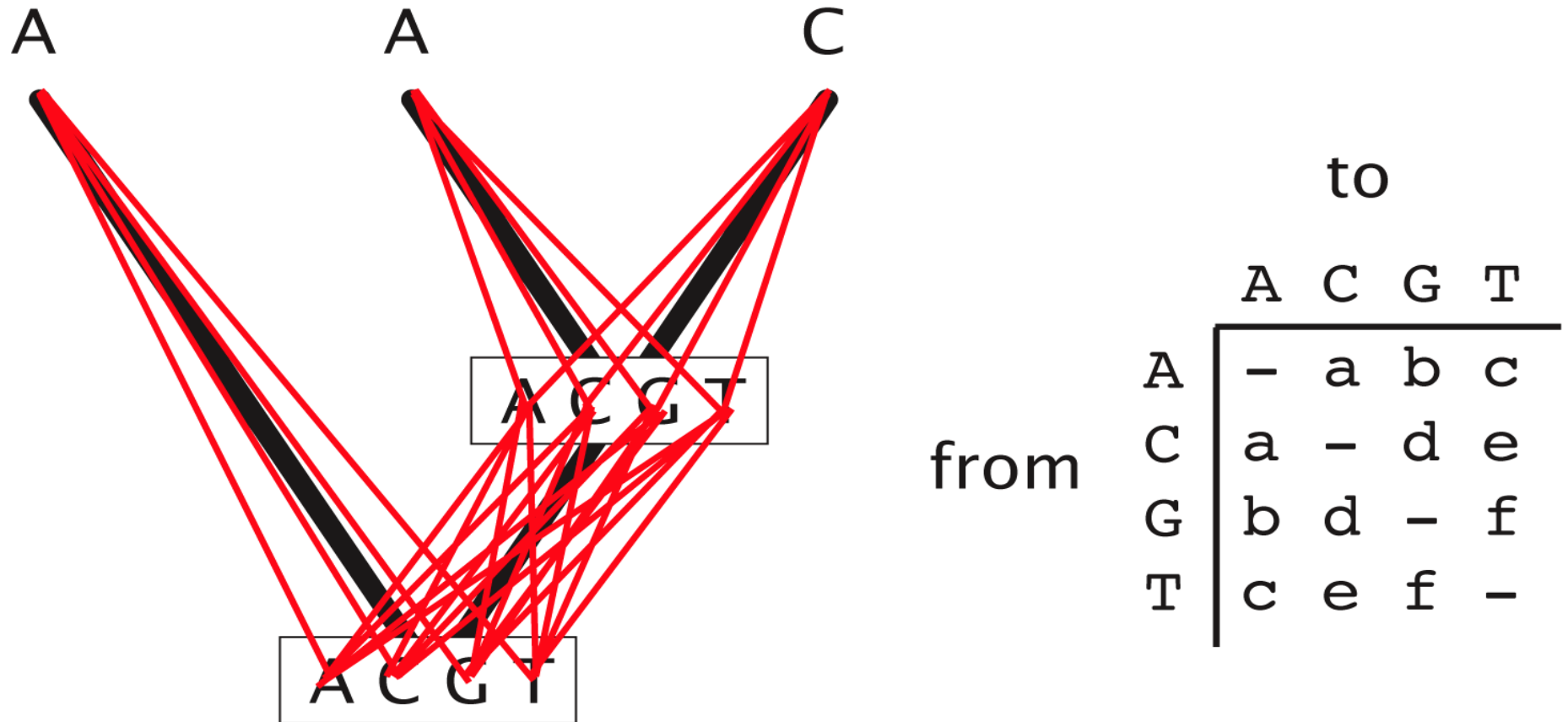


from

to

	A	C	G	T
A	-	a	b	c
C	a	-	d	e
G	b	d	-	f
T	c	e	f	-

Maximum likelihood



Likelihood = sum of all possible scenarios

Maximum likelihood

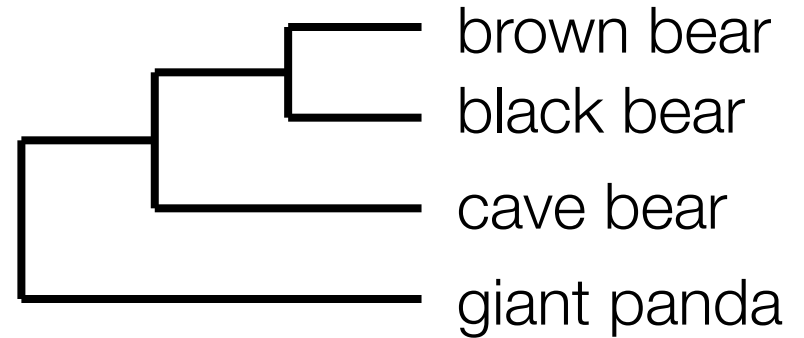
Likelihood is multiplied across sites

	L_1	L_2	L_3	...
brown bear	C	G	T	T A G T A C A C T
cave bear	C	G	A	T A G T T C A C T
black bear	C	G	T	A G T T T A C C
giant panda	C	A	T	T G G T T T A C T

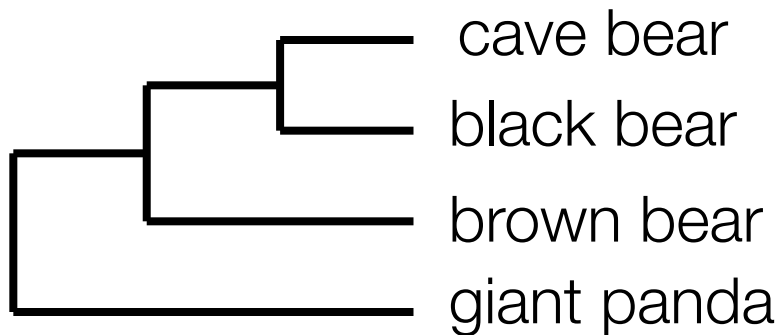
Likelihood values are very small!

Maximum likelihood

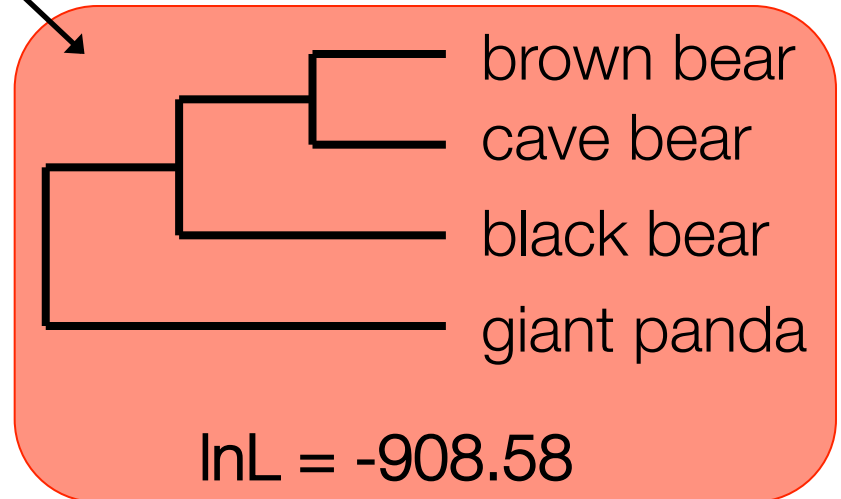
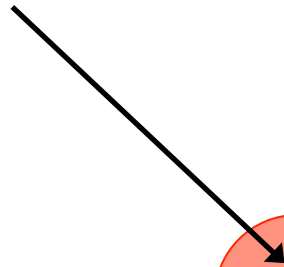
brown bear **C**G**T**T**A**G**T****A**C**A**C**T**
cave bear **C**G**A**T**A**G**T**T**C**A**C**T
black bear **C**G**T**T**A**G**T**T**T**A**C**C
giant panda **C**A**T**T**G**G**T**T**T**A**C**T



$$\ln L = -1203.83$$



$$\ln L = -1241.47$$

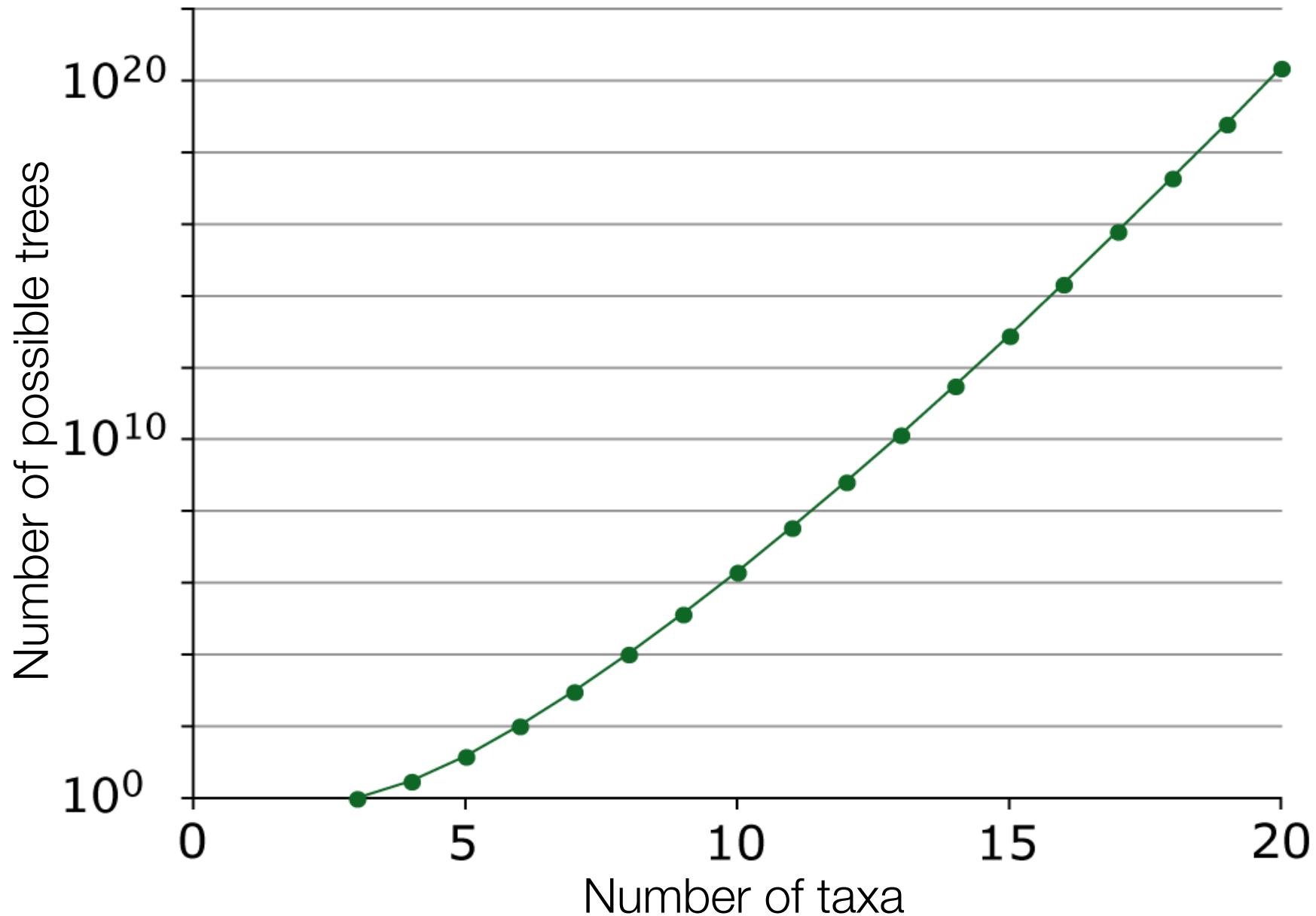


$$\ln L = -908.58$$

Likelihood optimisation

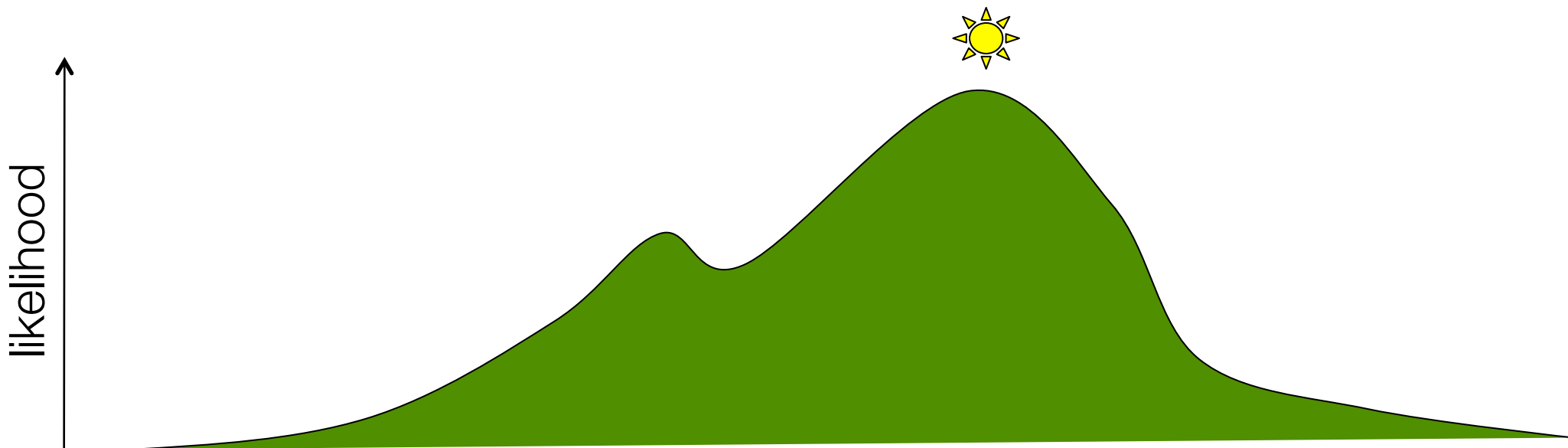
- Search through the space of possible trees and parameter values
- Calculate the likelihood for these
- Find best tree and model parameter values
- Multivariate optimisation

Searching tree space



Heuristic search

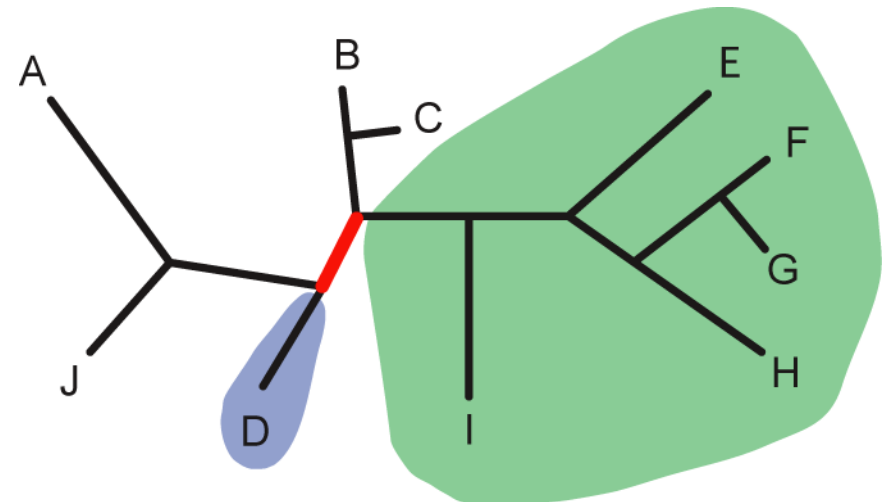
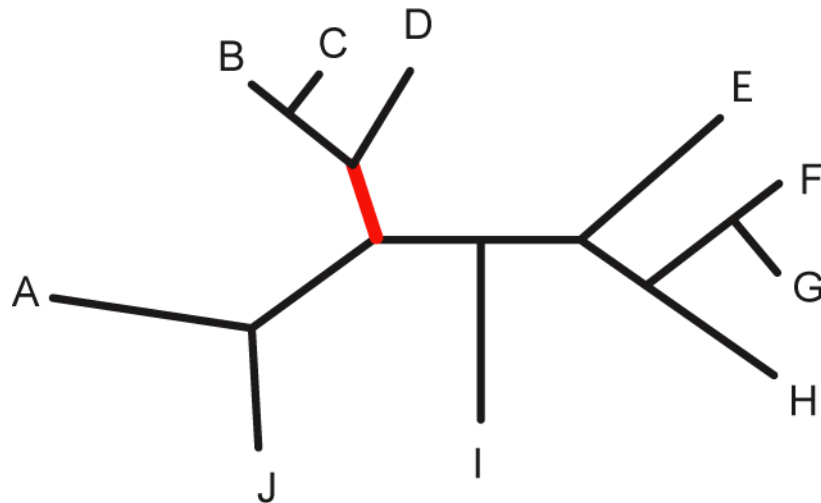
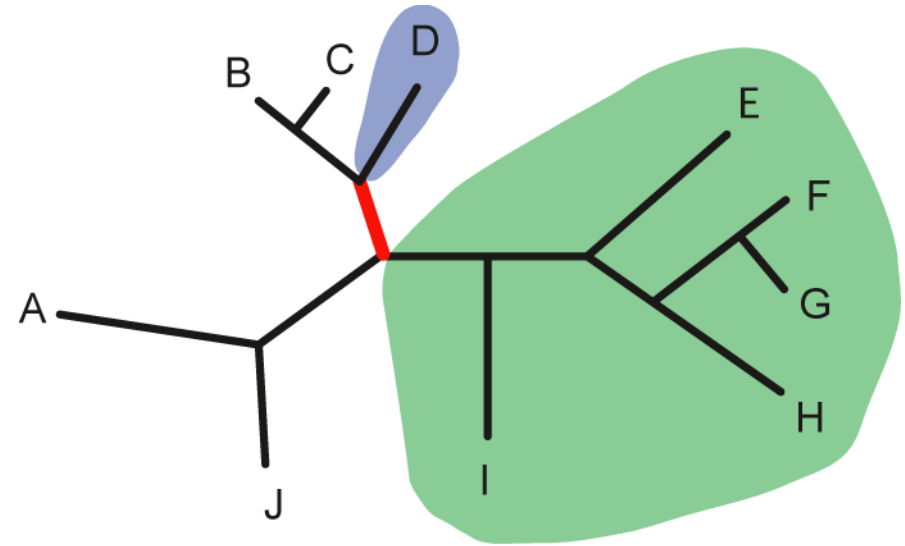
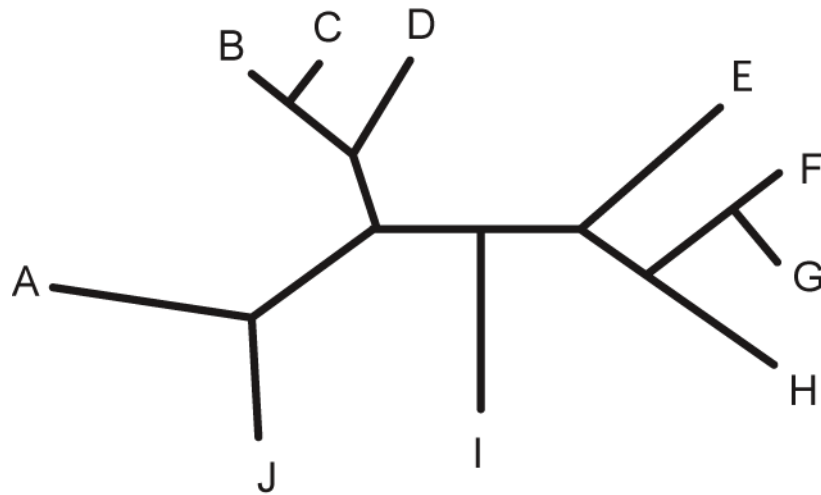
Heuristic search algorithms



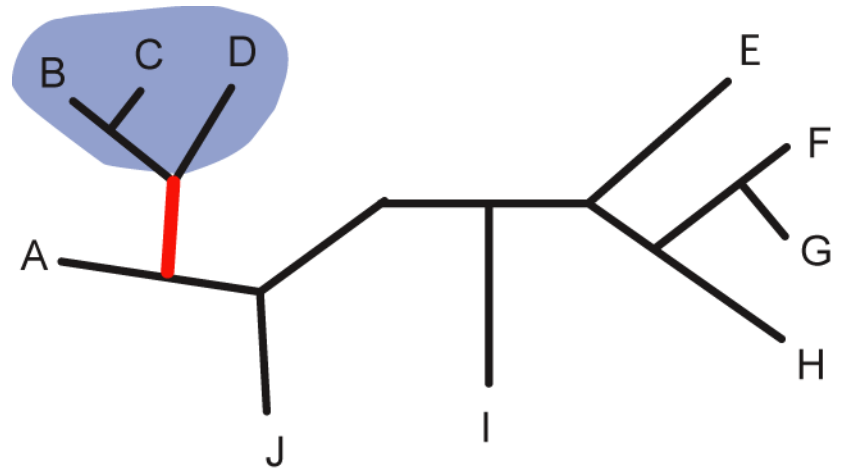
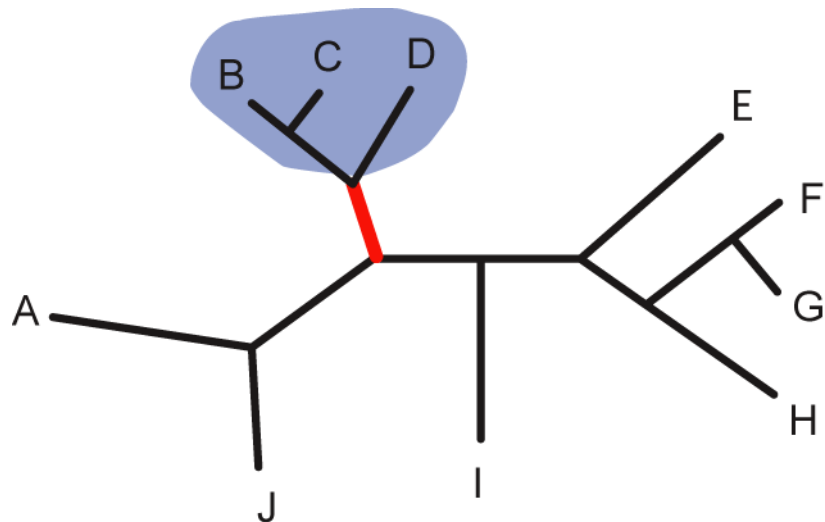
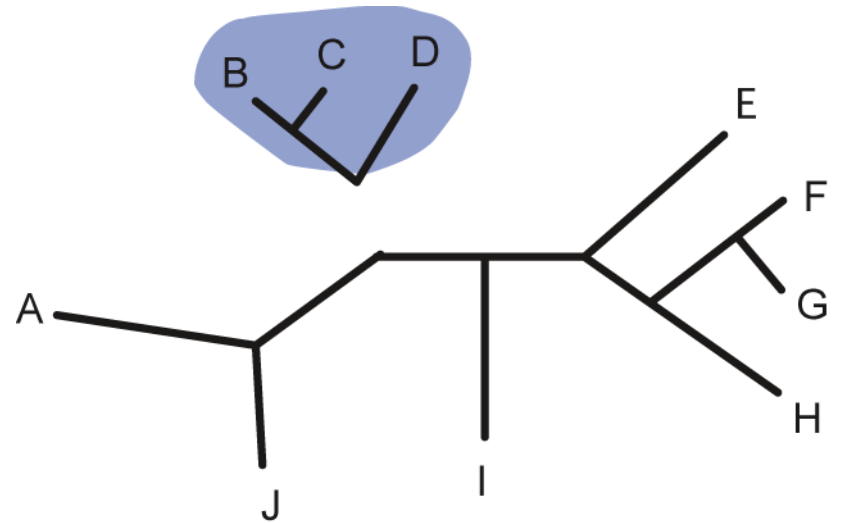
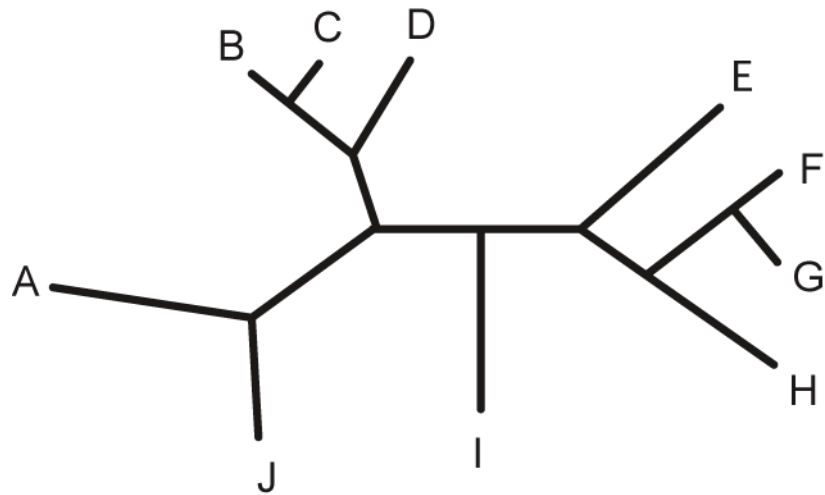
Methods of proposing changes to trees that result in nearby trees:

- Nearest-neighbour interchange (NNI)
- Subtree prune and regraft (SPR)
- Tree bisection and reconnection (TBR)

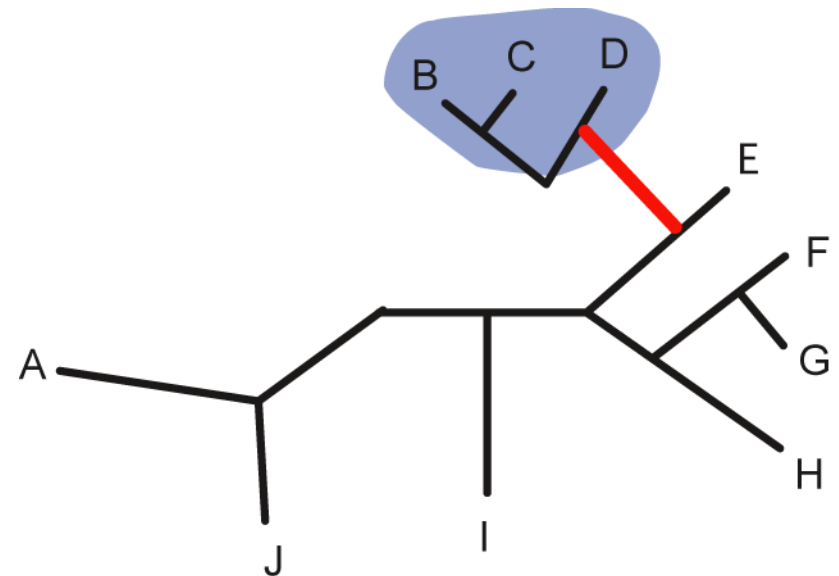
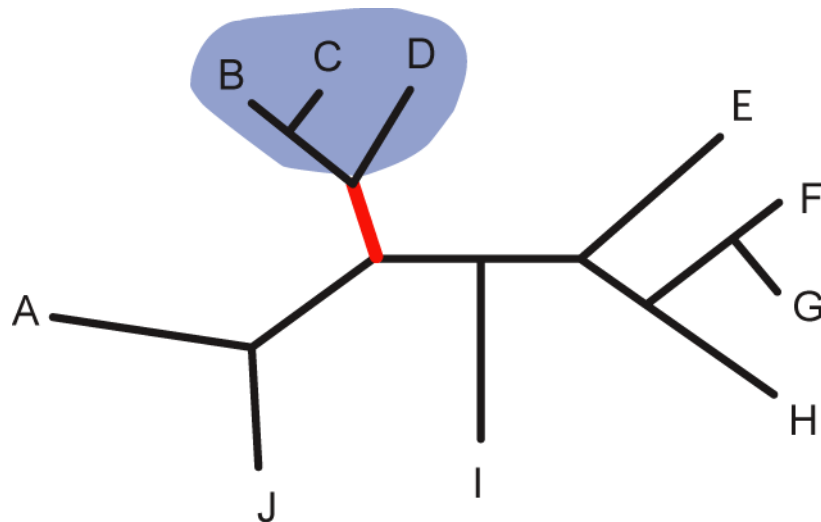
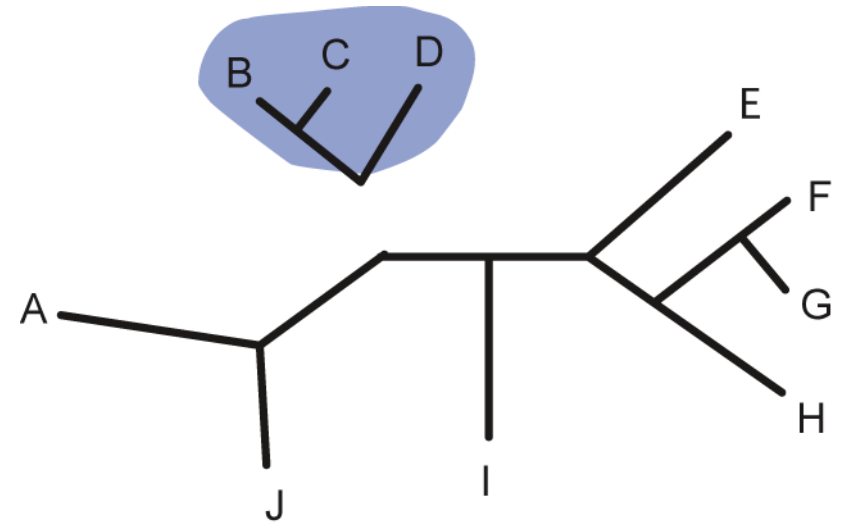
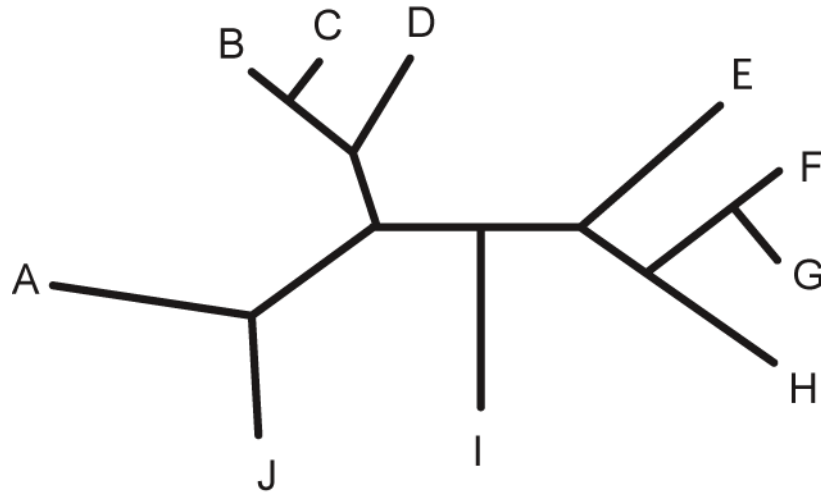
Nearest-neighbour interchange (NNI)



Subtree prune and regraft (SPR)

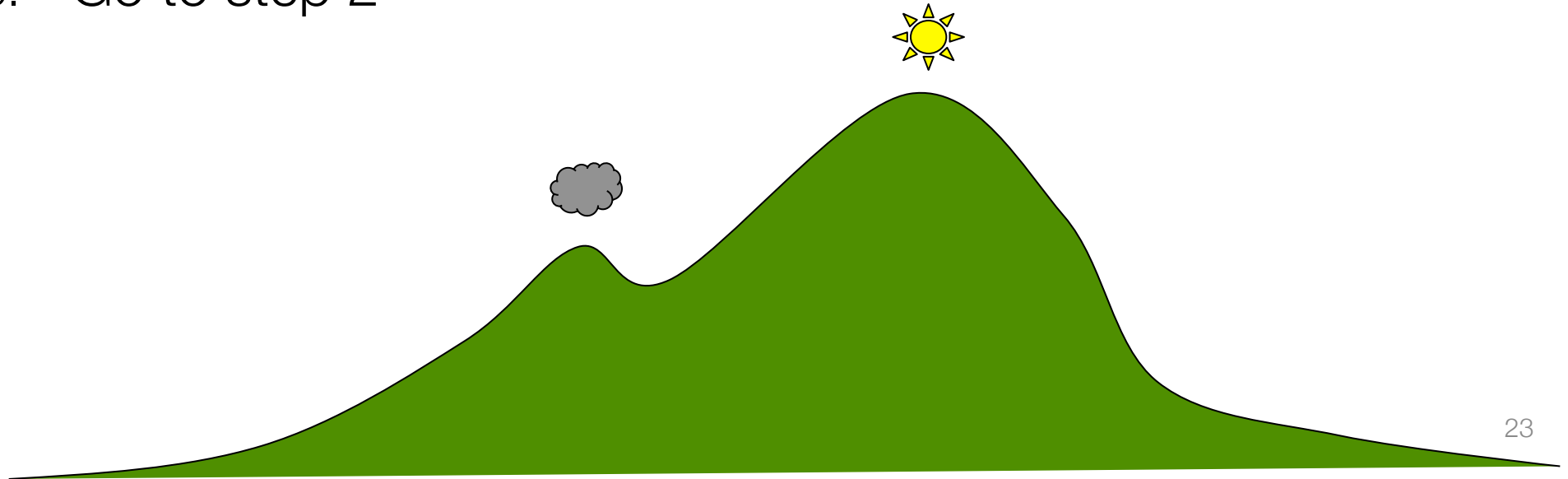


Tree bisection and reconnection (TBR)



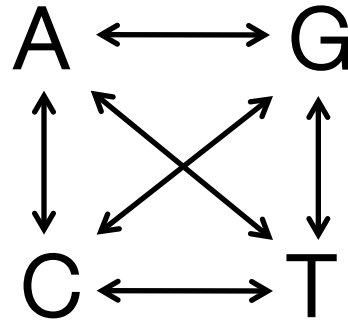
Heuristic search

1. Pick a starting tree (e.g., NJ or a random tree)
2. Use heuristic search to improve model parameters
3. Use heuristic search to improve branch lengths
4. Use NNI, SPR, and/or TBR to look for a better tree
5. Go to step 2

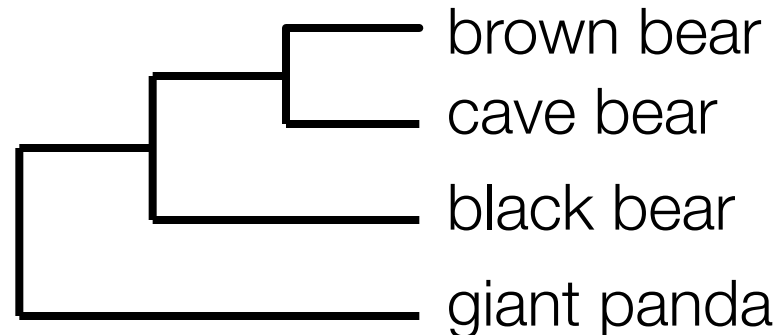


The result

A single set of maximum-likelihood estimates of model parameters



A single maximum-likelihood tree



Confidence intervals

- For MLEs of model parameters:
 - Can use the normal approximation (assumes symmetric variance around MLE)
 - 95% confidence interval is:
 $\text{MLE} \pm (1.96 \times \text{stdevMLE})$
- We cannot construct a confidence interval for the tree
 - Instead, uncertainty is estimated indirectly using **bootstrapping analysis**

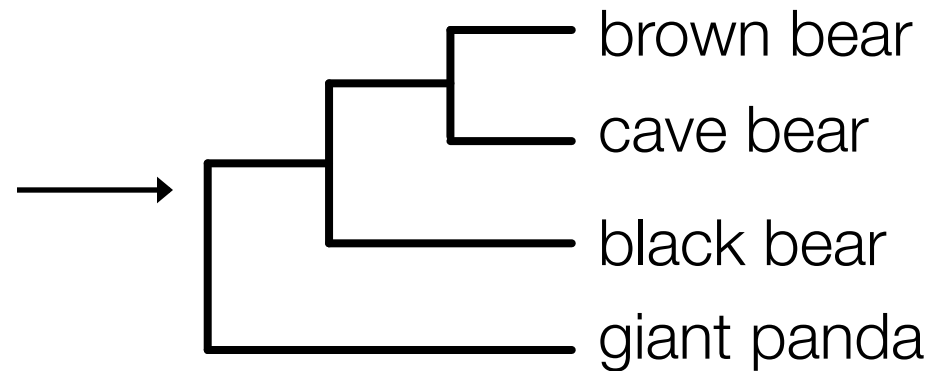
Bootstrapping

brown bear	CGTTAGTACACT
cave bear	CGATAGTTCACT
black bear	CGTTAGTTTACC
giant panda	CATTGGTTTACT

Repeat 1,000 times

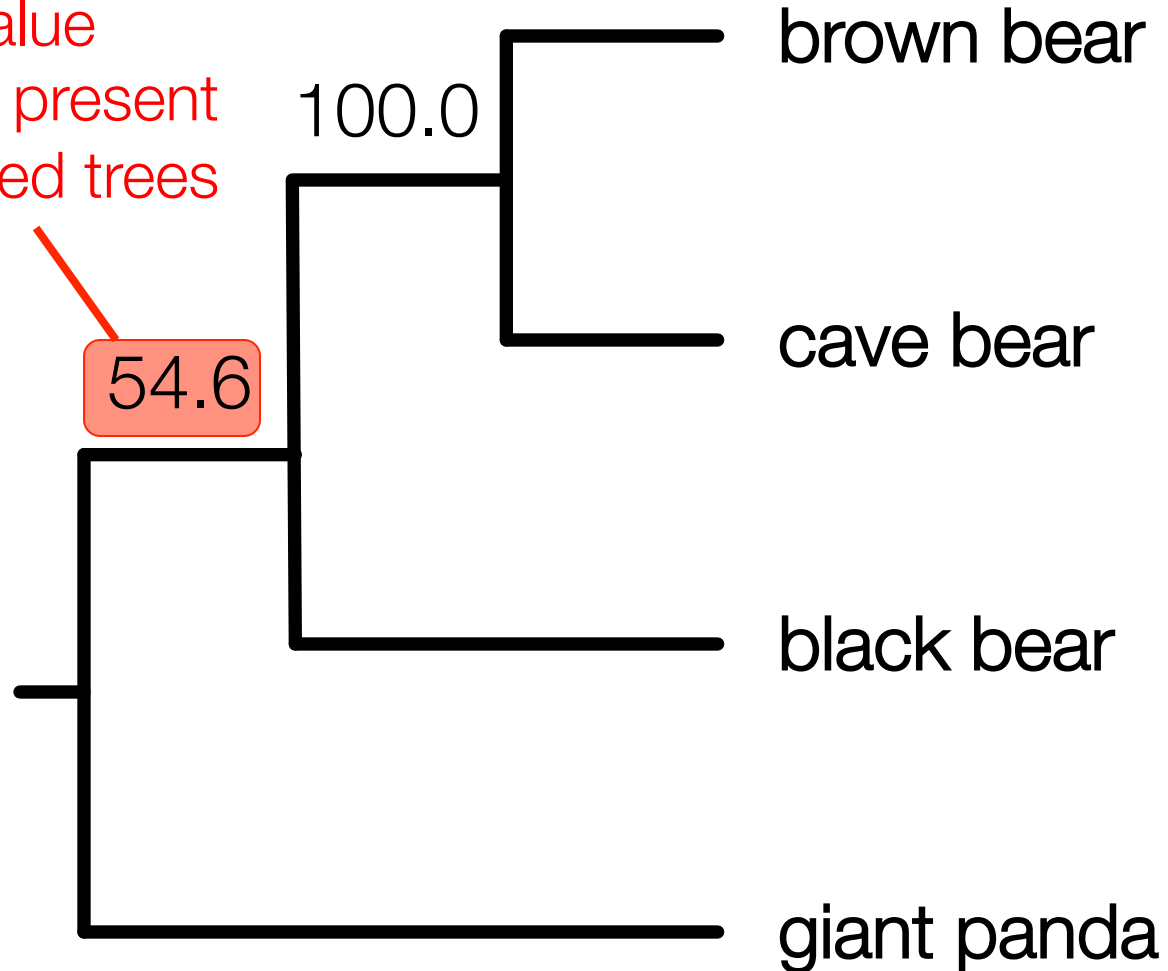
Pseudoreplication

brown bear	ATTACTGTCCCT
cave bear	ATTACTGTCCCA
black bear	ATCACTGTTCCCT
giant panda	GTTGCTATTCCCT



Bootstrapping

This value
is only present
in rooted trees



Topology tests

- Kishino-Hasegawa (KH) test
 - Test statistic: Difference in log-likelihood between two candidate trees
 - Problem: selection bias
- Shimodaira-Hasegawa (SH) test
 - Corrects for the selection bias in the KH test
 - Very conservative test
- Approximately-unbiased (AU) test
 - Less conservative than the SH test

Strengths and weaknesses

- **Strengths**
 - Rigorous statistical method
 - Desirable statistical properties
 - Highly robust to violations of assumptions
- **Weaknesses**
 - Not feasible to implement very parameter-rich models
 - Searching tree-space can be difficult
 - Need to rely on heuristic search methods
 - Bootstrapping analysis is very slow

Software

PHYLIP



PhyML



PAUP



Garli

MEGA



RAxML

Phylogenetic methods

	Algorithm-based	Optimality criterion	Other
No explicit substitution model		Maximum parsimony	
Explicit substitution model	Distance-based methods	Maximum likelihood	Bayesian inference

Go to Practical 1b: Model selection in MEGA

Go to Practical 1c: Maximum likelihood in
PhyML