

20.109  
Laboratory Fundamentals in  
Biological Engineering

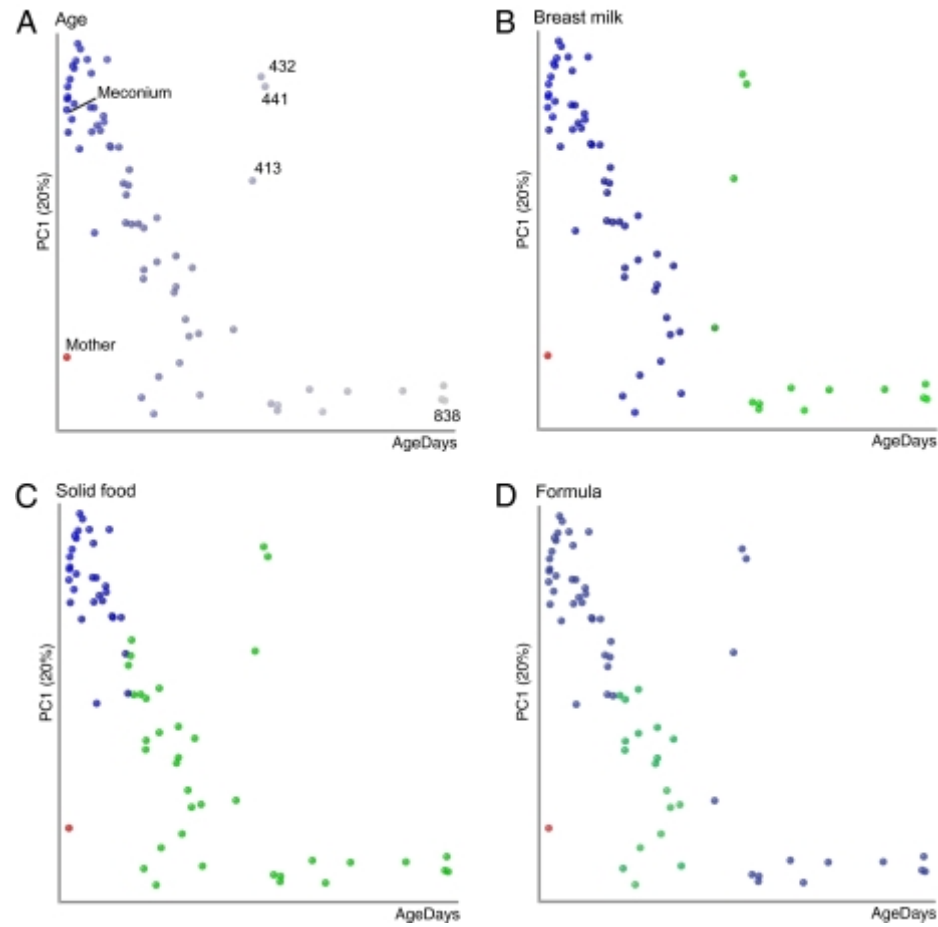
Module 1  
Nucleic Acid Engineering  
Lecture 7

# How do you compare the composition of two microbial communities?

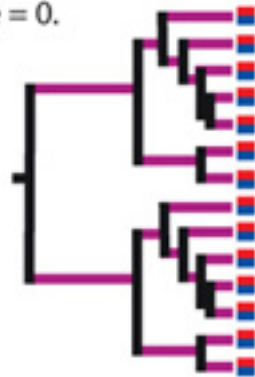
## Diversity

- $\alpha$  diversity: taxa within a sample
- $\beta$  diversity: between sample comparisons

# UniFrac



A. Identical sequence sets: all seqs in red + blue set. 100% branch length shared (purple).  
UniFrac score = 0.

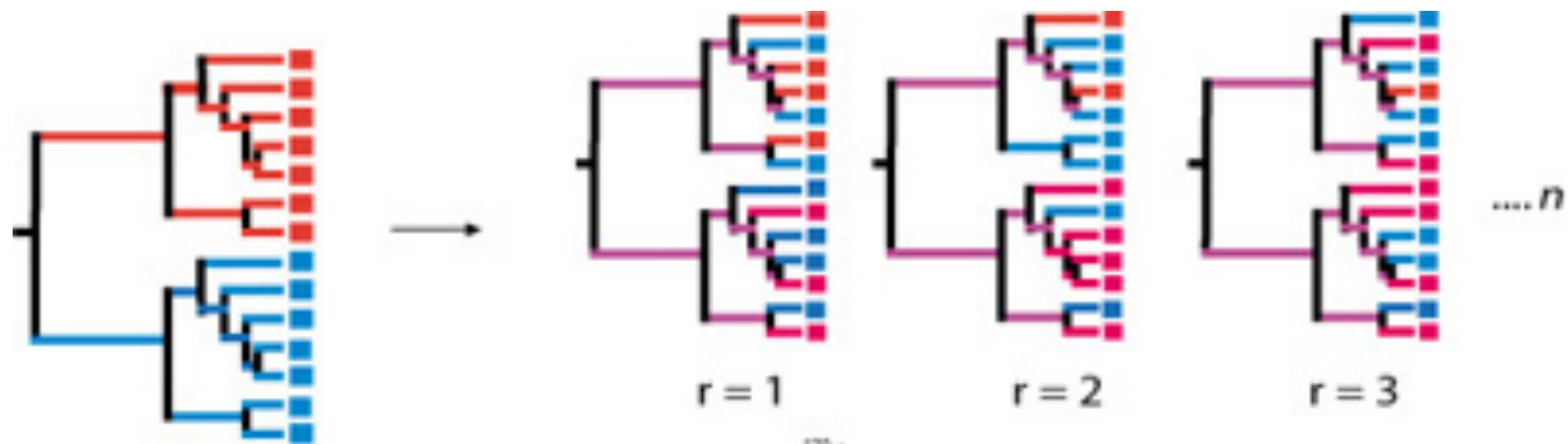


B. Related sequence sets: seqs in red have relatives in blue. ~50% branch length shared.  
UniFrac score = 0.5.

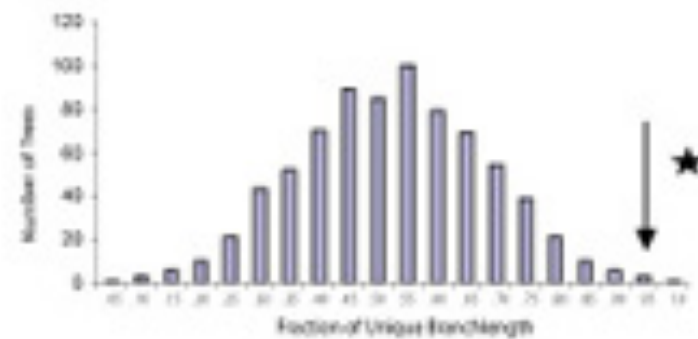


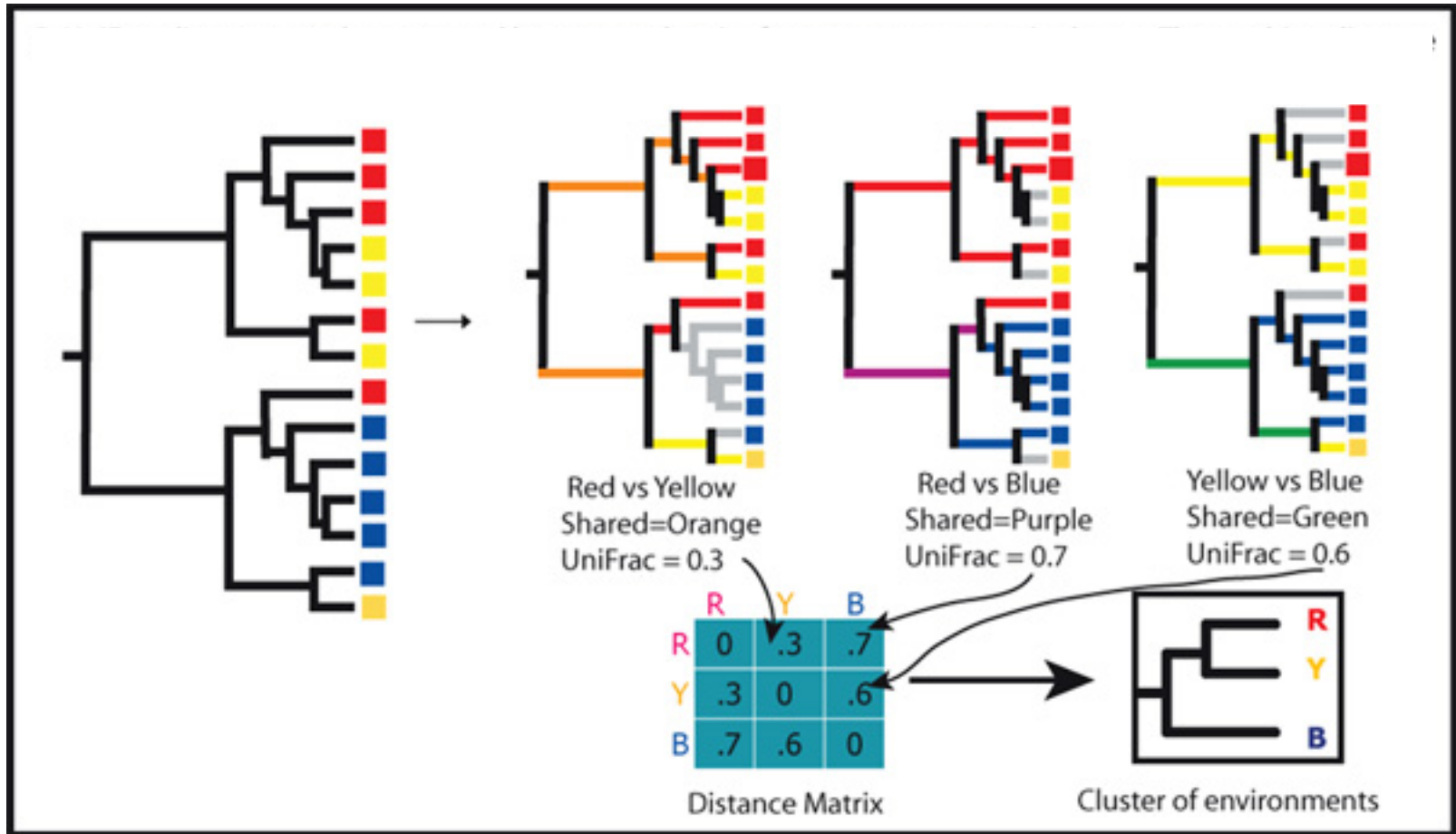
C. Unrelated sequence sets: seqs in red have no close relatives in blue. 0% branch length shared.  
UniFrac score = 1.



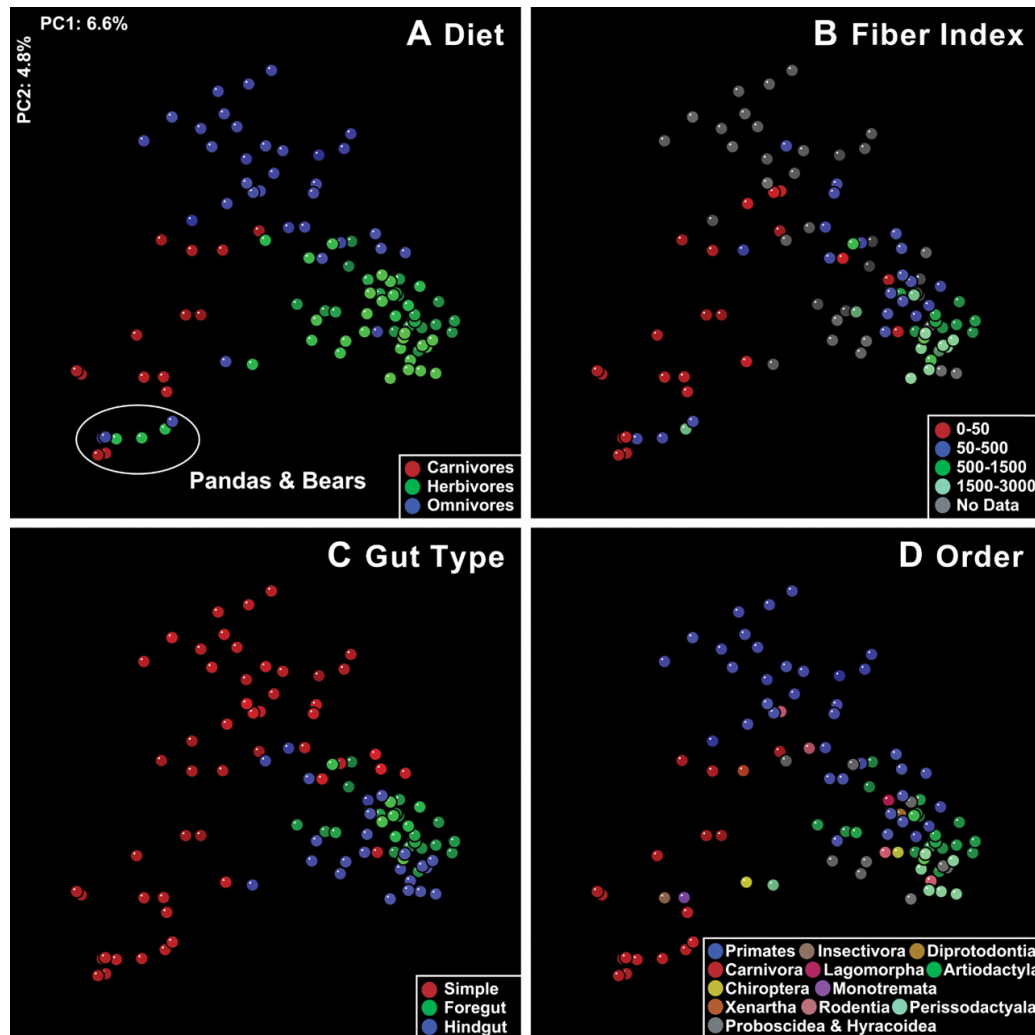


Observed





# Mammalian fecal bacterial communities clustered using PCoA



# UniFrac





# How to build a phylogenetic tree

1

- Collect data i.e. DNA

2

- Retrieve homologous sequences

3

- Multiple sequence alignment

4

- Model selection

5

- Assessing confidence in topology

# Model selection

- Scoring a matrix can't tell you which traits are derived and which are ancestral
- Need trees to infer evolutionary relationships
- Choose the **simplest** or **most likely** tree corresponding to the matrix



# Simplest vs most likely

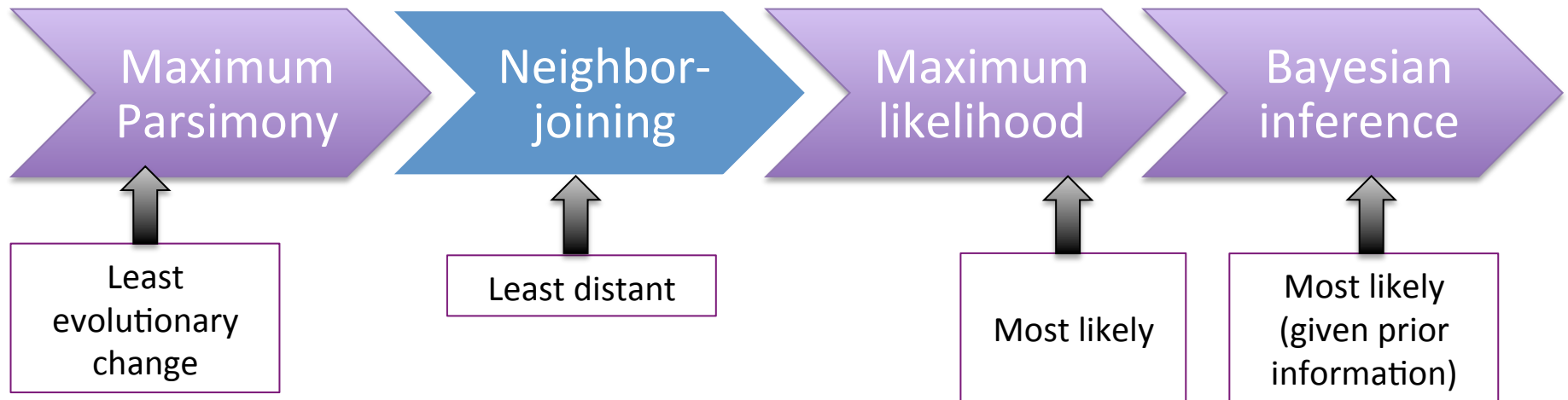
- We need a **metric** to decide which trees are better and which trees are worse
- **Optimality criterion** = a metric of quality (i.e. tree length, parsimony or likelihood) used to assess the **optimal tree**

# Optimality criteria

Which methods use an optimality criteria to decide on best tree?

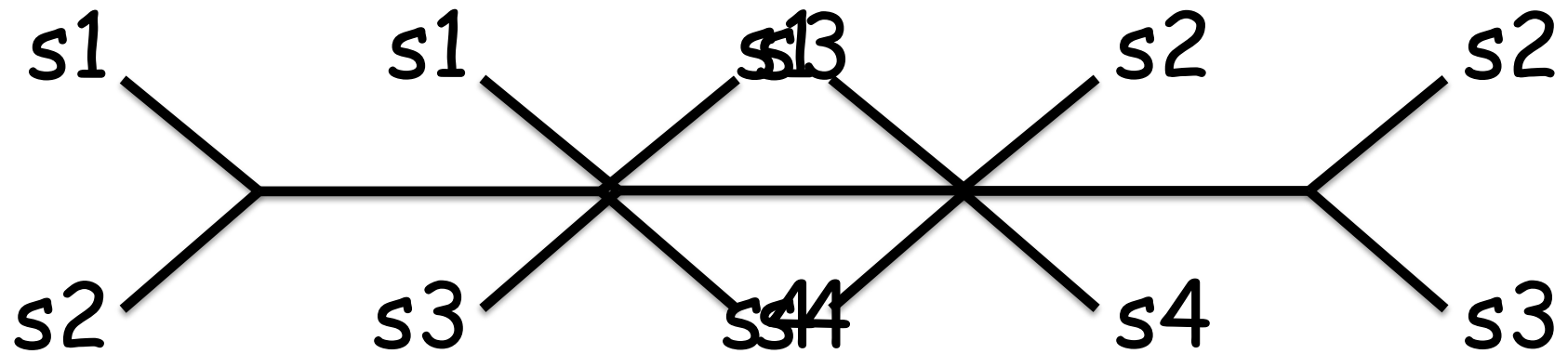


# Optimality criteria



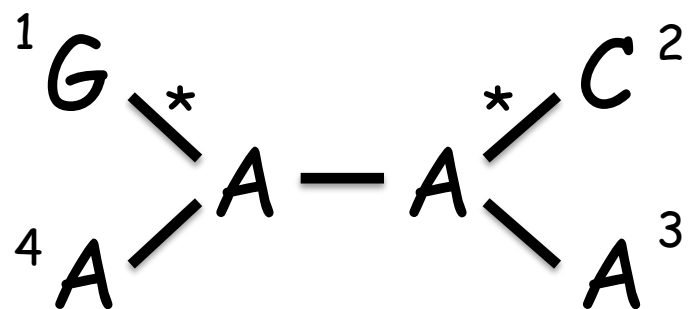
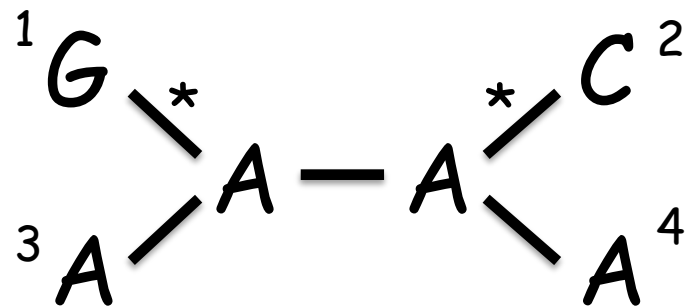
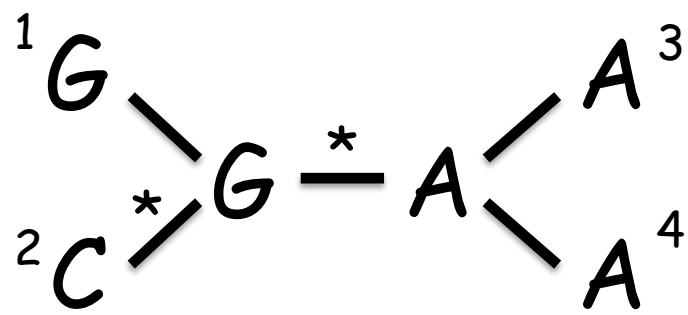
# Parsimony

S1	A	A	G	A	G	T	G	C	A
S2	A	G	C	C	G	T	G	C	G
S3	A	G	A	T	A	T	C	C	A
S4	A	G	A	G	A	T	C	C	G



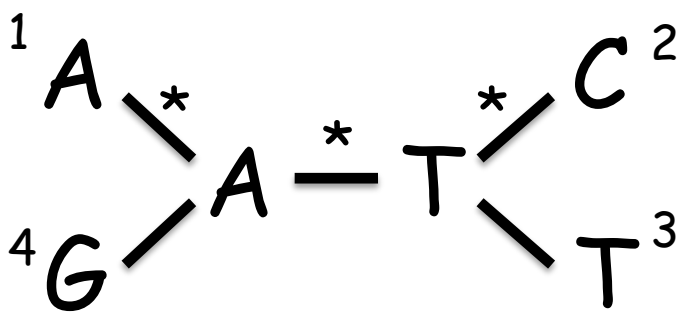
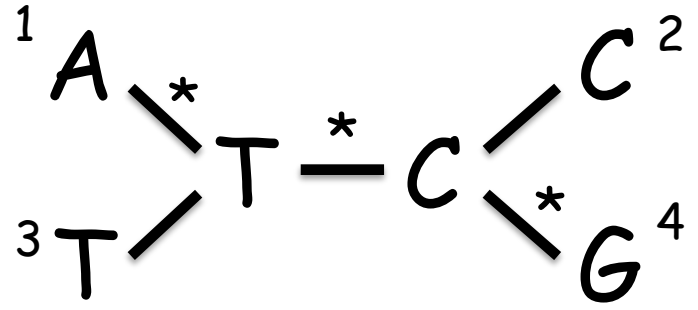
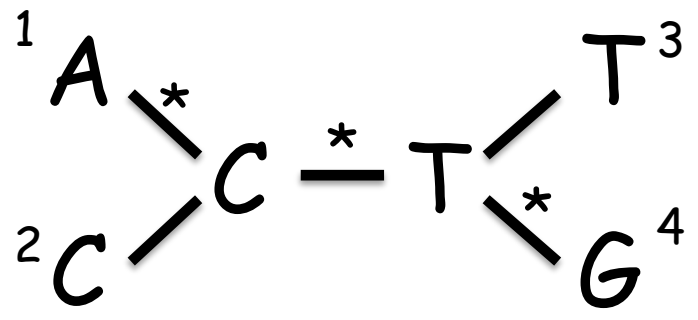
Site 3

G  
C  
A  
A



Site 4

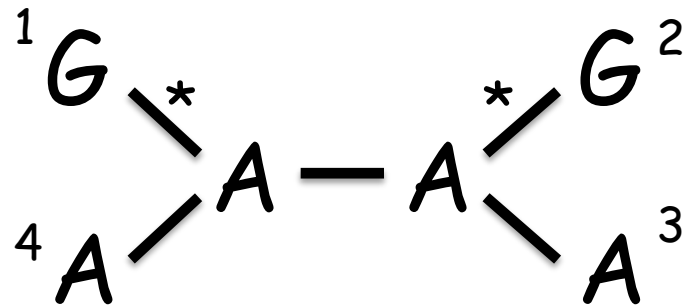
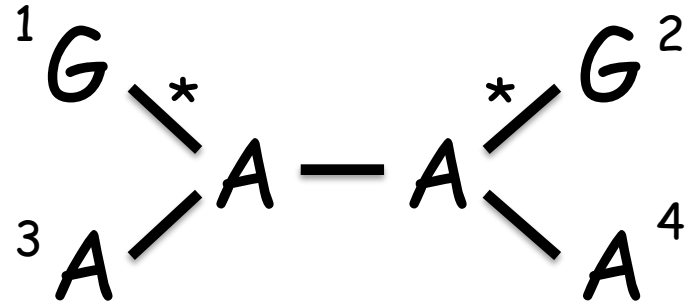
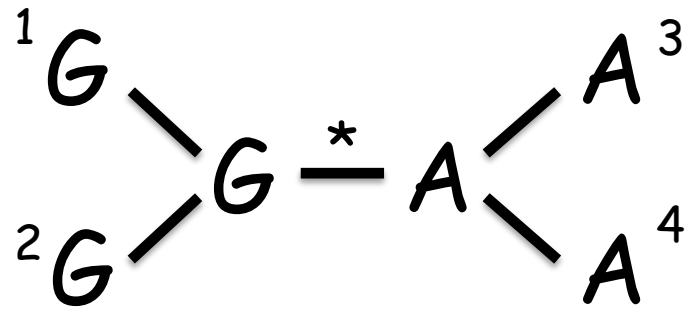
A  
C  
T  
G



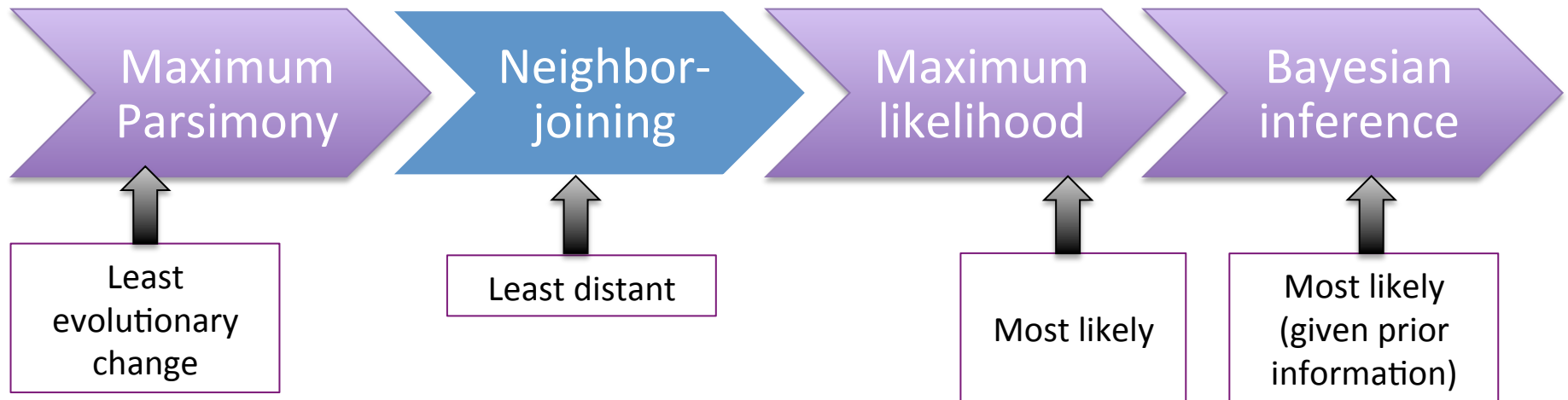


Site 5

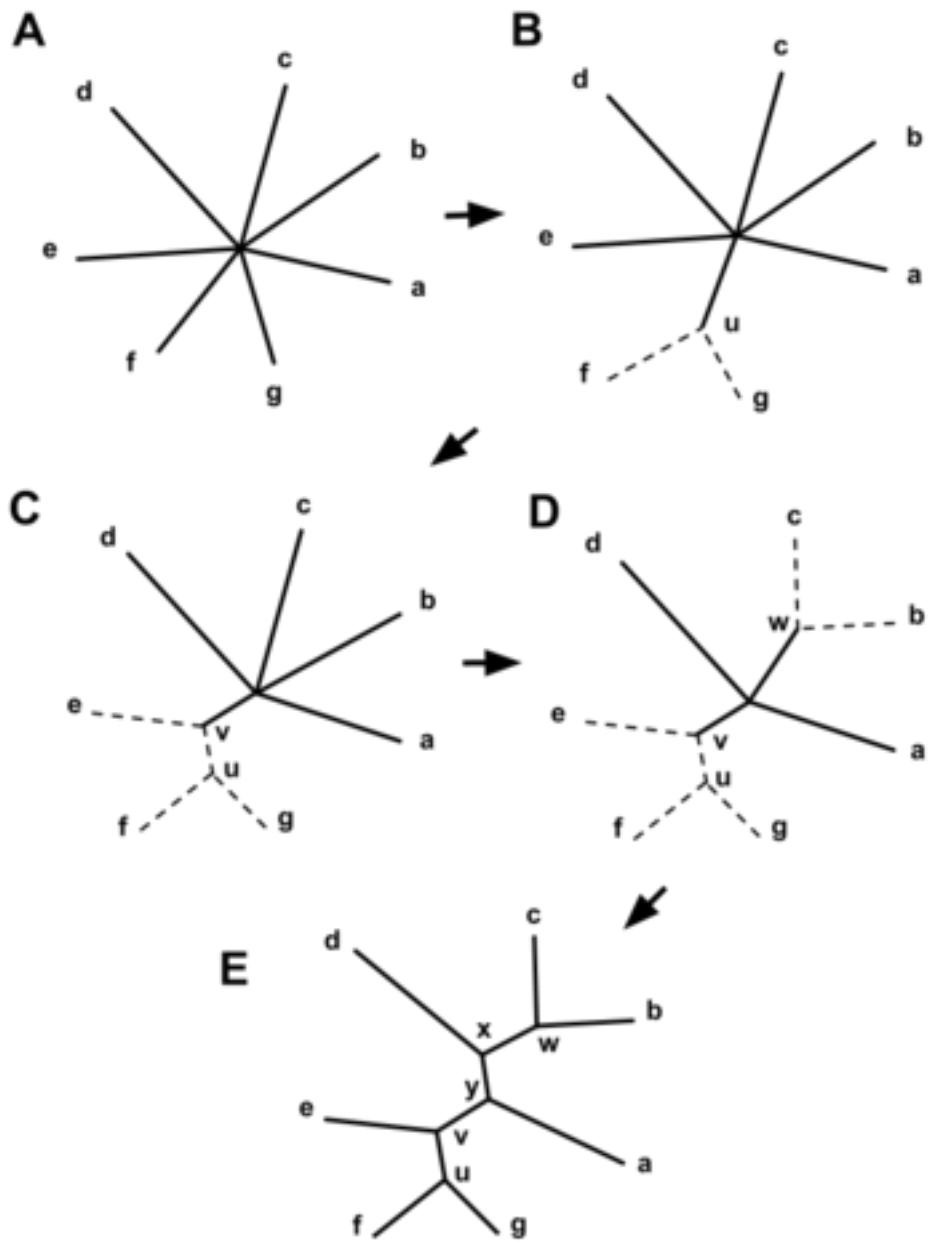
G  
G  
A  
A



# Optimality criteria



NJ



# Optimality criteria

