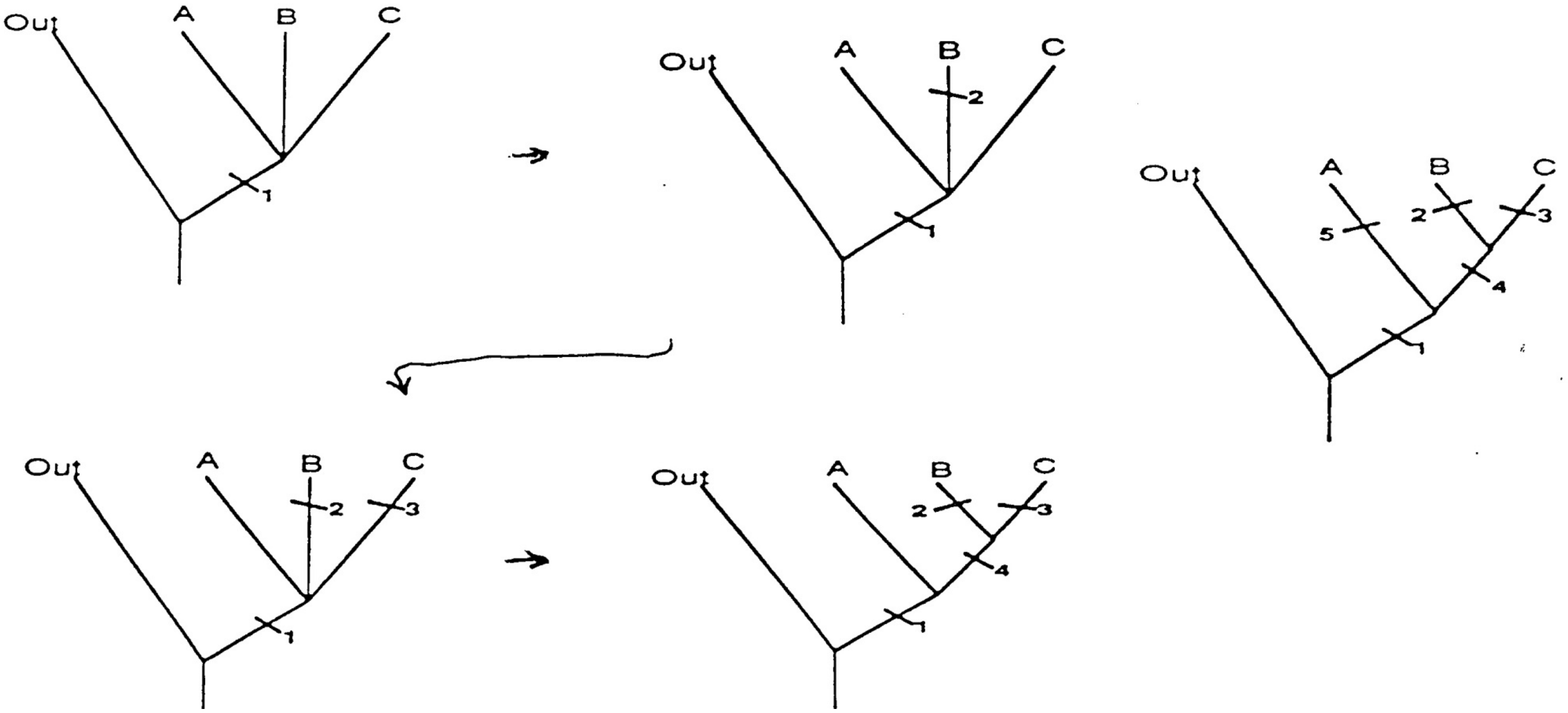
Hennigs method

	characters (transformation series)				
	1	2	3	4	5
outgroup	0	0	0	0	0
A	1	0	0	0	1
B	1	1	0	1	0
C	1	0	1	1	0



Wagner algorithm

First we have to introduce few terms (or definitions):

1) A particular character (X) of a particular taxon (A) is defined as $X(A,i)$, where i is the ith character in a vector of i characters.
2) The vector of characters for a particular taxon is defined as $\Sigma X(A,i)$. For example, the character vector for M is:

$$\Sigma X(A,i) = 1\ 1\ 0\ 0\ 0\ 0.$$

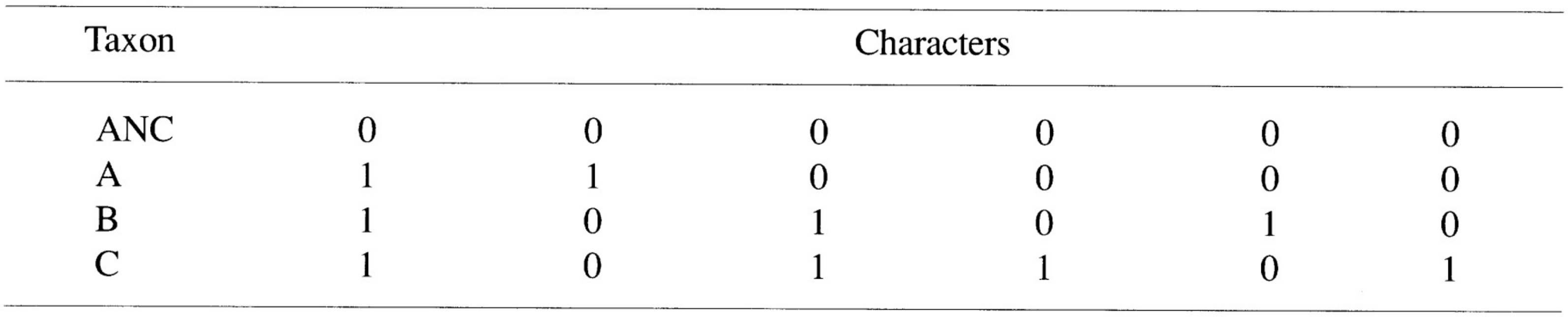
3) The difference (D) between two taxa is the sum of the absolute differences between their characters:

$$D(C,B) = \Sigma \left| X(C,i) - X(B,i) \right| .$$

We calculate this in the following manner :

$$D(C,B) = \Sigma \left| X(C,i) - X(B,i) \right| =$$

$$= (1 - 1) + (0 - 0) + (1 - 1) + (1 - 0) + (0 - 1) + (1 - 0) = 3$$

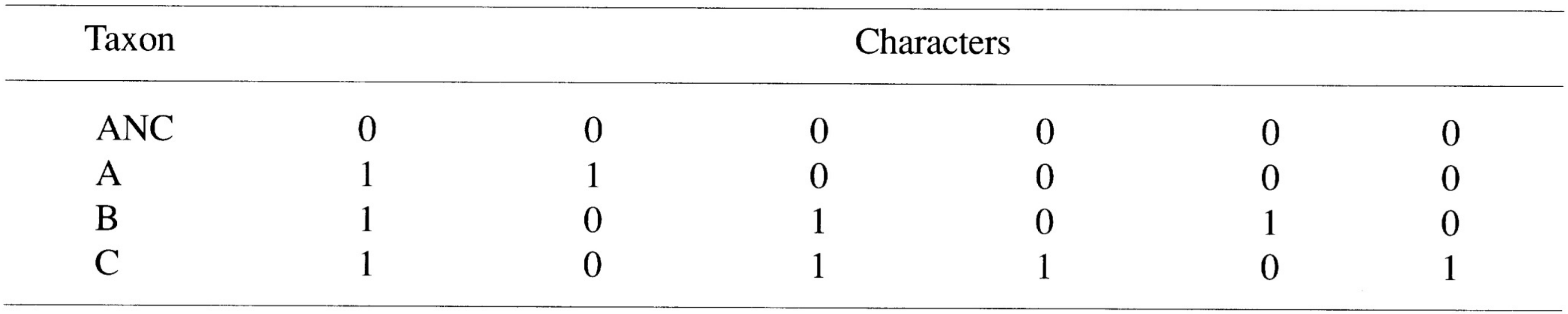


Taxon	Characters					
ANC	0	0	0	0	0	0
A	1	1	0	0	0	0
B	1	0	1	0	1	0
C	1	0	1	1	0	1

4) The interval (INT) of a taxon is the length of the line between that taxon and its ancestor. For example, the interval of B is:

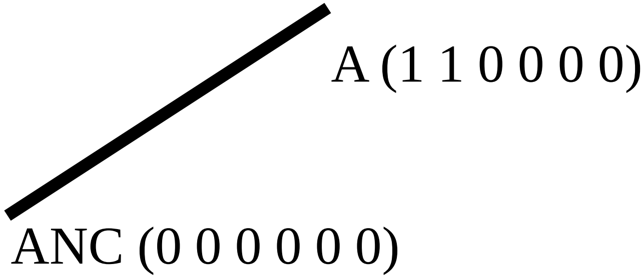
$$INT(B) = D[B, ANC(B)]$$

where INT(B) is the interval of taxon B, ANC(B) is the hypothetical ancestor of B, and D[B,ANC(B)] is the path length distance of B to its ancestor.



Taxon	Characters					
ANC	0	0	0	0	0	0
A	1	1	0	0	0	0
B	1	0	1	0	1	0
C	1	0	1	1	0	1

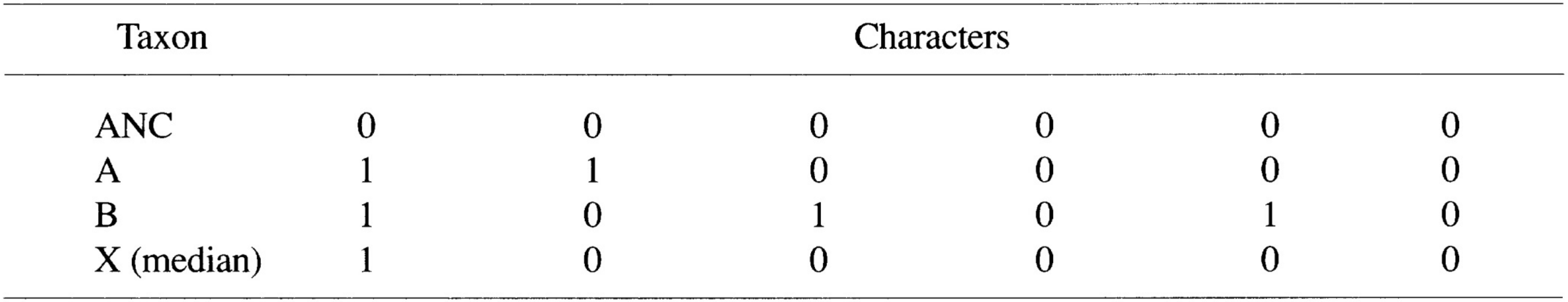
1. ANC vector in table represents the outgroup.
2. We calculate difference D from ANC for each taxon:
   - D(A,ANC)=2
   - D(B,ANC)=3
   - D(C,ANC)=4.
3. We create interval for this taxon. A is closest (in difference) to ANC, so we will build the interval INT(A,ANC):
   - INT(A,ANC) = D(A,ANC) = 2.



A (1 1 0 0 0 0)

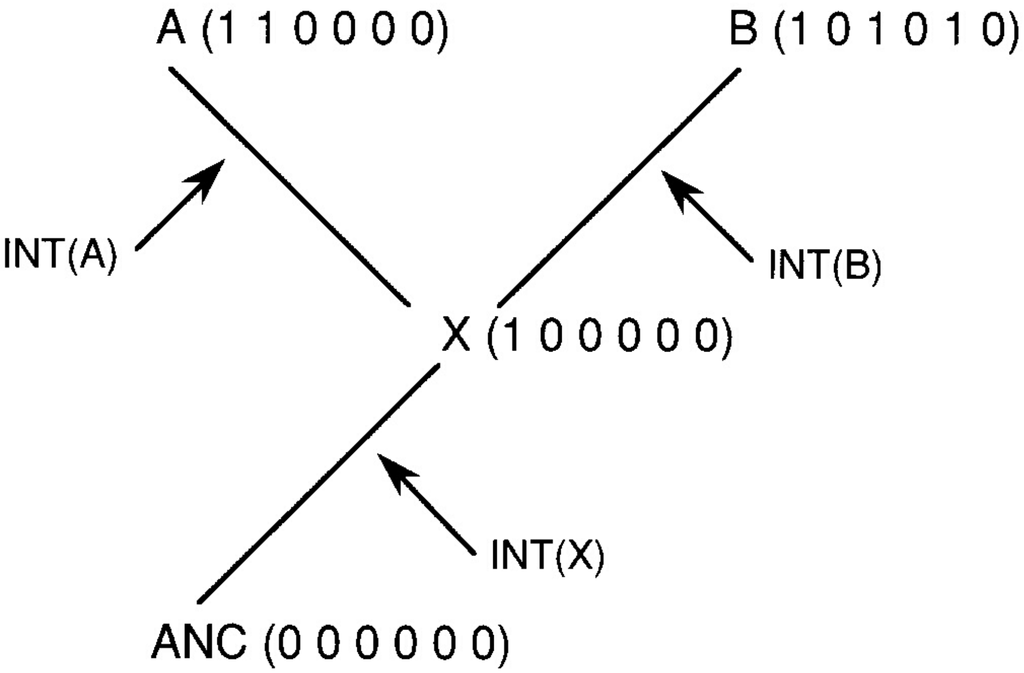
ANC (0 0 0 0 0 0)

4. We select taxon, that has smallest difference D to ANC. It is taxon B.
5. Then we will look for the interval, that has smallest difference D to the taxon B. Currently there is only one interval INT(A,ANC), so there is no choice. So we do not need to calculate D[B,INT(A)].
6. We connect B to INT(A), by constructing hypothetical ancestor (X), characters of which are medians of transformation series ANC, A and B, so those three taxa that we included until now.



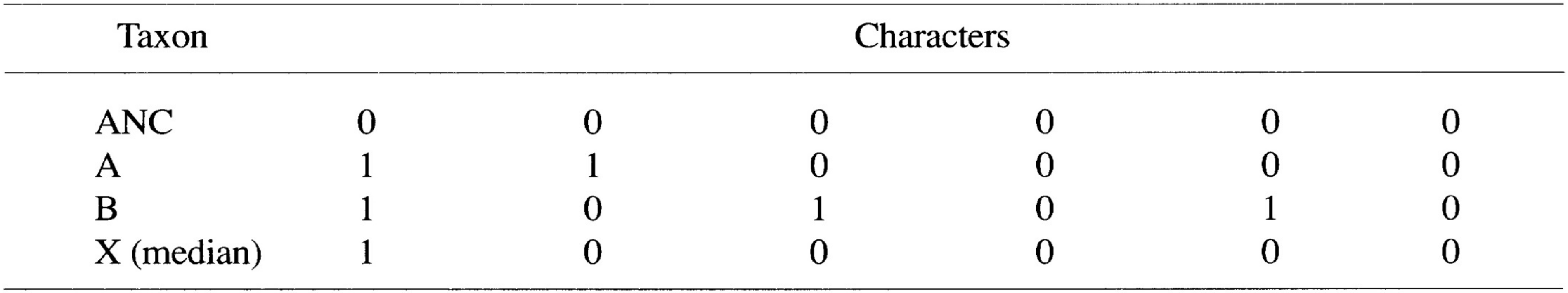
Taxon	Characters					
ANC	0	0	0	0	0	0
A	1	1	0	0	0	0
B	1	0	1	0	1	0
X (median)	1	0	0	0	0	0

Taxon	Characters					
ANC	0	0	0	0	0	0
A	1	1	0	0	0	0
B	1	0	1	0	1	0
X (median)	1	0	0	0	0	0



7. Going back to the step 4 of algorithm, we add further taxon, which has smallest D to ANC. There is only one remaining taxon, C. According to algorithm we should connect it with the interval that has smallest difference to C. Therefore, we must calculate three interval difference values, one for each interval in the tree. For this we have to calculate differences between taxa, their ancestors and taxon C (we already know the difference between C and ANC):

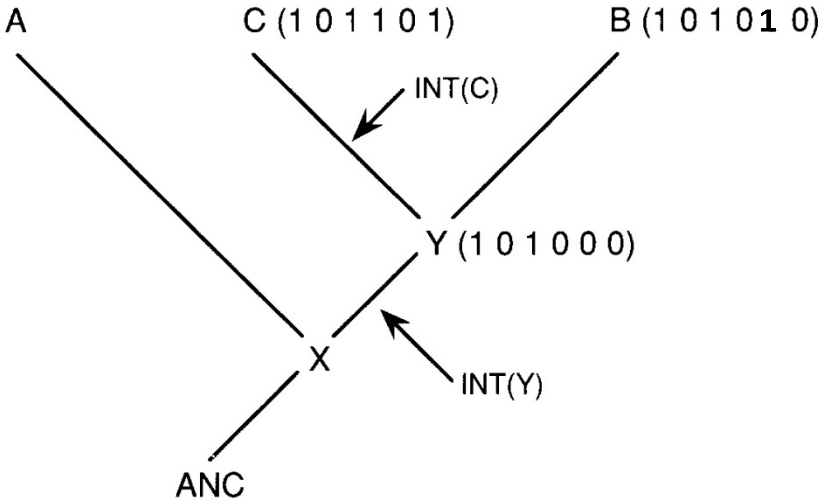
$$D(A,X) = \left| X(A,i) - X(X,i) \right| = 1$$
$$D(B,X) = \left| X(B,i) - X(X,i) \right| = 2$$
$$D(C,X) = \left| X(C,i) - X(X,i) \right| = 3$$
$$D(C,A) = \left| X(C,i) - X(A,i) \right| = 4$$
$$D(C,B) = \left| X(C,i) - X(B,i) \right| = 3$$
$$D(X,ANC) = \left| X(X,i) - X(ANC,i) \right| = 1$$

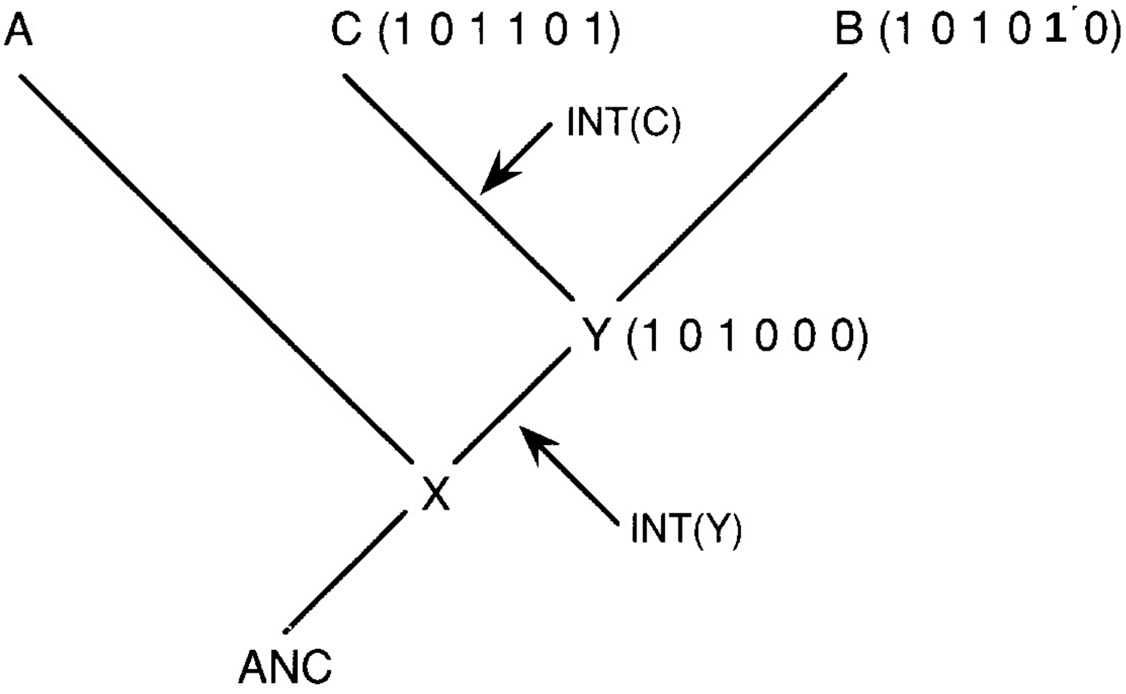


Taxon	Characters					
ANC	0	0	0	0	0	0
A	1	1	0	0	0	0
B	1	0	1	0	1	0
X (median)	1	0	0	0	0	0

So we can calculate differences to intervals:
D[C,INT(A)] = D(C,A) + D(C,X) - D(A,X)/2 = (4 + 3 - 1)/2 = 3
D[C,INT(B)] = D(C,B) + D(C,X) - D(B,X)/2 = (3 + 3 - 2)/2 = 2
D[C,INT(X)] = D(C,X) + D(C,ANC) - D(X,ANC)/2 = (3 + 4 - 1)/2 = 3

Because the difference between C and INT(B) has the smallest value, we construct another hypothetical ancestor (Y) and connect C to the tree through this new ancestor to INT(B). To calculate the character vector for this new ancestor, take the median of the vectors of the three appropriate taxa, X, B, and C. So we have the resulting tree:





Taxon	Characters					
X	1	0	0	0	0	0
B	1	0	1	0	1	0
C	1	0	1	1	0	1
Y (median)	1	0	1	0	0	0

Phylogenetic tree building methods

method of building trees	**type of data**	
	distances	**DNA sequences or other characters**
clustering algorithm	UPGMA neighbor-joining tree	
optimality criterion	minimum evolution tree	→ parsimony maximum likelihood Bayesian analysis

Willi Hennig (German edition 1950, Spanish edition 1965, English edition 1966, 1968)

Evolutionary relationships are inferred from SYNAPOMORPHIES

descendant states, similarity – based on common origin (**HOMOLOGY**)

Hennig did not provide any method for identification of synapomorphies.

Farris (1967, 1970, 1972, 1973, 1976, 1977, 1979, 1980, 1982, et al. 1970, and others)

Criterion for identification of synapomophies is PARSIMONY

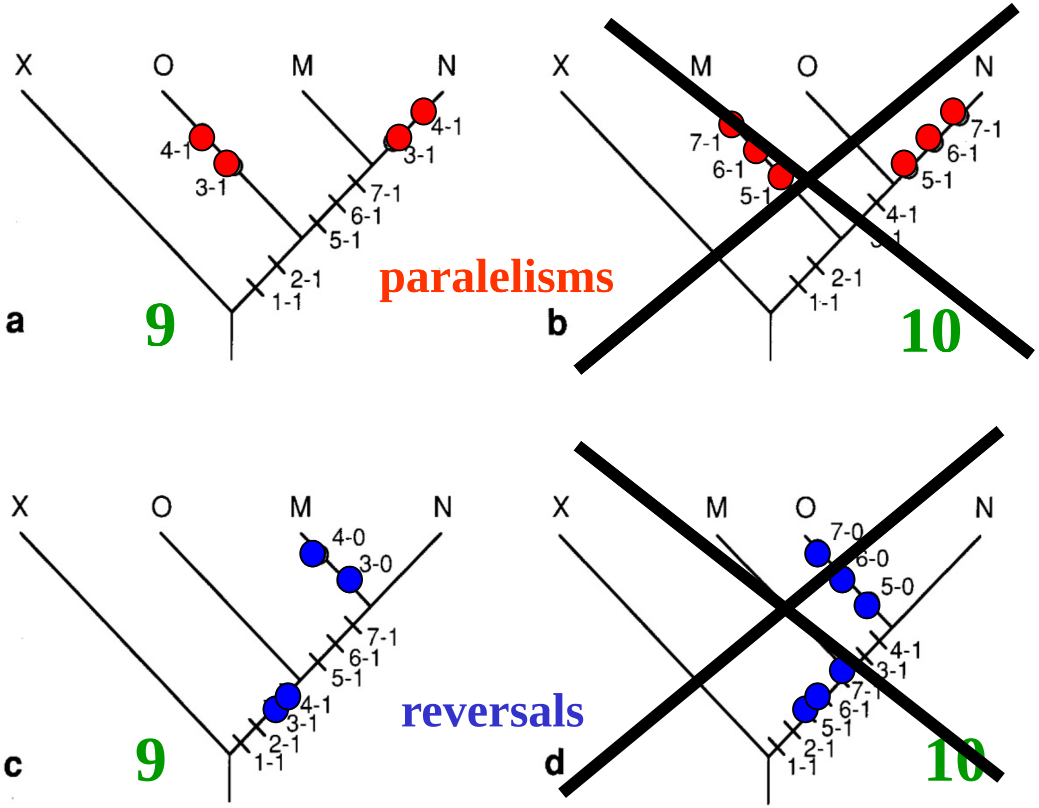
Nixon, WHS worshop 2006

Maximum parsimony (MP)

Simple, intuitive and logical method (based on Medieval logic, Occam's razor – prefers the simplest solutions), no statistics

Minimalizes ad hoc explanations – homoplasy

Maximizes explanatory value



9

paralelisms

10

9

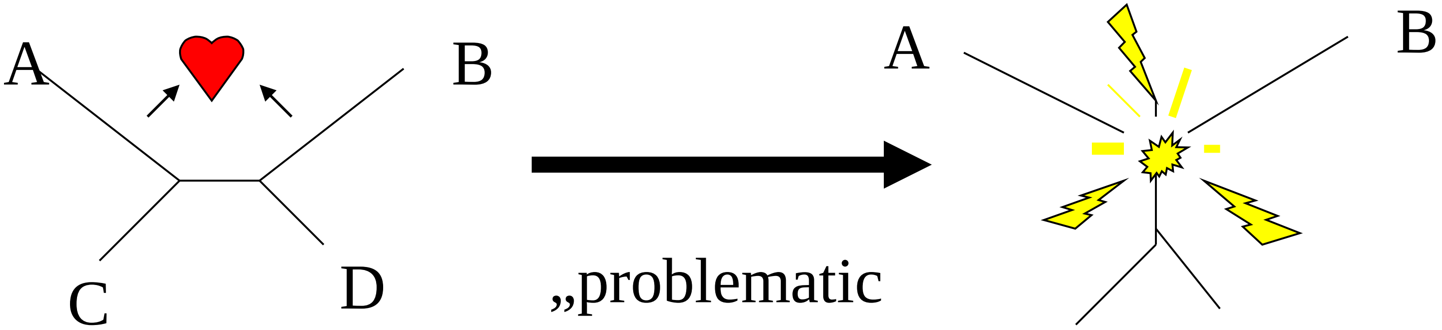
reversals

10

Parsimony can provide misleading results – long branch attraction

Felsenstein zone – some branches or internodes look shorter because of numerous mutations

A -> C -> A, etc.



„problematic affair“

Correct tree

Incorrect tree

Parsimony

1) **Wagner parsimony:** treats characters as ordered such that the change from one character to another implies change through any intervening characters in the transformation series. Characters are allowed to reverse freely.

2) **Fitch parsimony:** treats characters in a transformation series as unordered such that change from one character to another character within a transformation series does not imply changes through possible intervening characters. Characters are allowed to reverse freely.

3) **Dollo parsimony:** requires every synapomorphy to be uniquely derived, i.e., appearing only once on the tree. The synapomorphy may reverse, but once reversed it cannot reappear. But a “relaxed” Dollo criterion, which amounts to assigning a weight to the cost of reversal, might be more appropriate in some cases.

4) **Camin-Sokal parsimony:** makes the assumption that character evolution is irreversible. This is true, philosophically, because time is irreversible and the reversals are really new apormorphies. But we cannot know this a priori. This criterion is rarely used.

Simple parsimony: selection of the optimal cladogram depends only on the amount of the character changes on the tree, reversals and parallelisms have an equal value.

General parsimony: depends on more detailed knowledge of characters. Some characters might be treated as ordered, others as unerdered, in some cases certain characters might have higher weight.

Equally parsimonious trees:

1. equal topology, but different interpretation of characters – phylogenetic relationships are identical, resulting classification is the same.

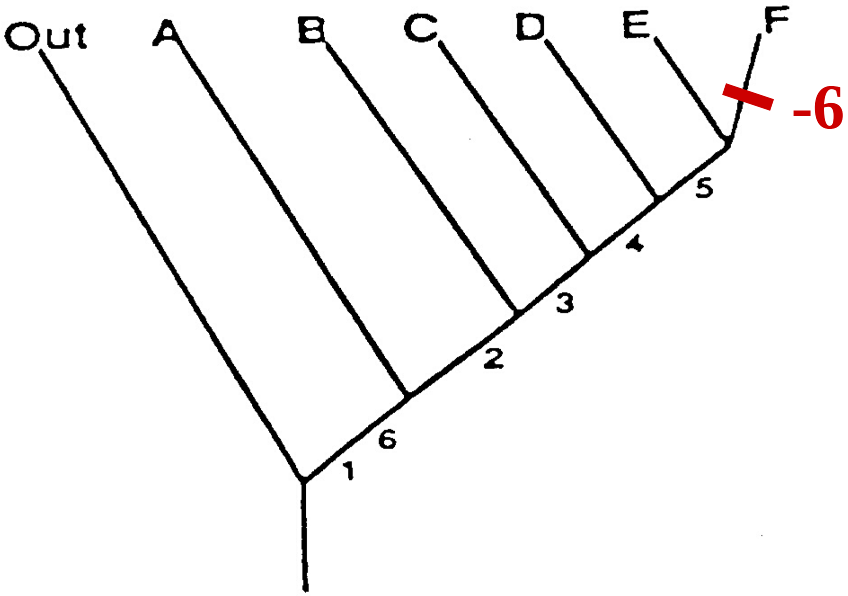
2. different topology – phylogenetic relationships are different, resulting classifications are different

Consistency index, CI, Kluge a Farris (1969) measures of how transformation series and entire data matrices “fit” particular tree topologies. The consistency index is the ratio of the minimum amount of changes (steps) it might show in the matrix (m) and the amount of change (steps) it does show on a particular tree (s) multiplied by 100: CI = m/s x 100

	transformation series					
	1	2	3	4	5	6
OG	0	0	0	0	0	0
taxon A	1	0	0	0	0	1
B	1	1	0	0	0	1
C	1	1	1	0	0	1
D	1	1	1	1	0	1
E	1	1	1	1	1	1
F	1	1	1	1	1	0

transf. series	changes
1	0 -> 1
2	0 -> 1
3	0 -> 1
4	0 -> 1
5	0 -> 1
6	0 -> 1

Total amount of changes of characters in the matrix is 6.



Summary of the character changes in the cladogram:

Transf. series	changes
1	0 -> 1
2	0 -> 1
3	0 -> 1
4	0 -> 1
5	0 -> 1
6	0 -> 1 -> 0

Total amount of changes (steps) in the cladogram is 7

then CI=6/7 x 100 = 85,7

Retention index, RI, Farris (1989) The retention index measures the fraction of apparent synapomorphy to actual synapomorphy:

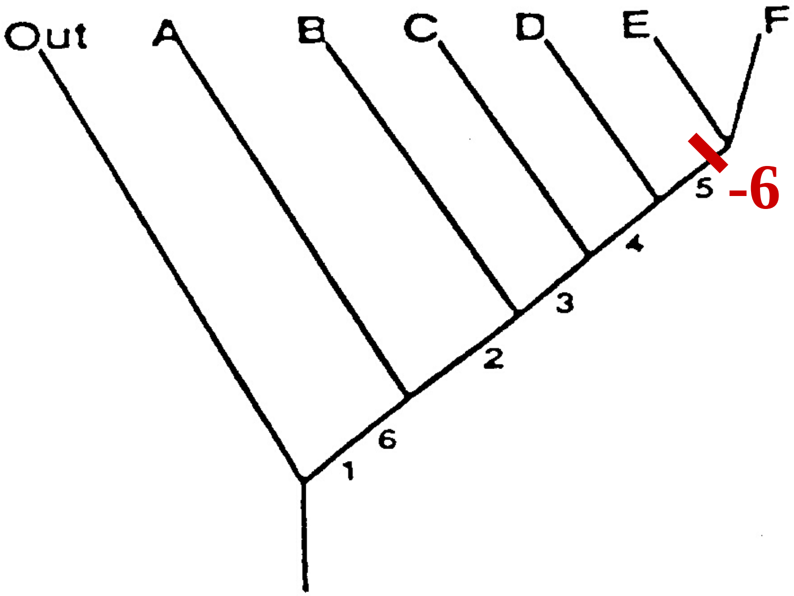
RI = (g - s) / (g - m) x 100

s, m defined the same way as for CI, g how many steps would it take to explain evolution within the transformation series under the worst possible condition, which is totoal amount of taxa with character 1 or 0 (whichever is smaller), summarized for all transformation series (resp. characters).

		transformation series				
	1	2	3	4	5	6
OG	0	0	0	0	0	0
taxon A	1	0	0	0	0	1
B	1	1	0	0	0	1
C	1	1	1	0	0	1
D	1	1	1	1	0	1
E	1	1	1	1	1	**0**
F	1	1	1	1	1	0

transf. series	changes
1	0 -> 1
2	0 -> 1
3	0 -> 1
4	0 -> 1
5	0 -> 1
6	0 -> 1

Total amount of changes of characters in the matrix is 6.



Summary of the character changes in the cladogram:

Transf. series	changes
1	0 -> 1
2	0 -> 1
3	0 -> 1
4	0 -> 1
5	0 -> 1
6	0 -> 1 -> 0

Total amount of changes (steps) in the cladogram is 7

then CI=6/7 x 100 = 85,7

CI index is equal, even if cladogram is different

	transformation series					
	1	2	3	4	5	6
OG	0	0	0	0	0	0
taxon A	1	0	0	0	0	1
B	1	1	0	0	0	1
C	1	1	1	0	0	1
D	1	1	1	1	0	1
E	1	1	1	1	1	1
F	1	1	1	1	1	0

transf. series	changes
1	0 -> 1
2	0 -> 1
3	0 -> 1
4	0 -> 1
5	0 -> 1
6	0 -> 1

Total amount od changes of characters in the matrix is 6. Number of changes on the cladogram is 7.

Transf. series	Max. amount of steps
1	1
2	2
3	3
4	3
5	2
6	2

RI = (13 - 7) / (13 - 6) x 100 = 85,7

Total maximum amount of steps is 13

	transformation series					
	1	2	3	4	5	6
OG	0	0	0	0	0	0
taxon A	1	0	0	0	0	1
B	1	1	0	0	0	1
C	1	1	1	0	0	1
D	1	1	1	1	0	1
E	1	1	1	1	1	0
F	1	1	1	1	1	0

transf. series	changes
1	0 -> 1
2	0 -> 1
3	0 -> 1
4	0 -> 1
5	0 -> 1
6	0 -> 1

Total amount od changes of characters in the matrix is 6. Number of changes on the cladogram is 7.

Transf. series	Max. amount of steps
1	1
2	2
3	3
4	3
5	2
6	3

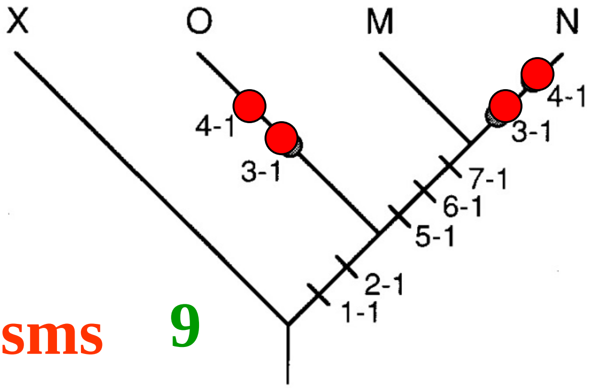
RI = (14 - 7) / (14 - 6) x 100 = 87,5

Total maximum amount of steps is 14

Instead of 85,7

Optimizing trees

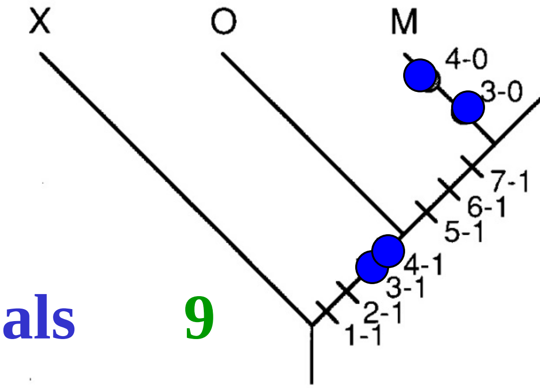
a posteriori **activity that does not help to construct trees but evaluates already constructed trees. In the case that we have more equally parsimonious trees of the same topology, taking into account the some features of characters we may select tree where we prefer parallelisms or reversals.**



parallelisms **9**

DELTRAN

DELays evolutionary TRANSformation of a character on a tree



reversals **9**

ACCTRAN

ACCelerates evolutionary TRANsformation of a character on a tree

Tree building methods

Exact methods

Exhaustive search, implicit enumeration
Branch-and-bound

Heuristic algorithms

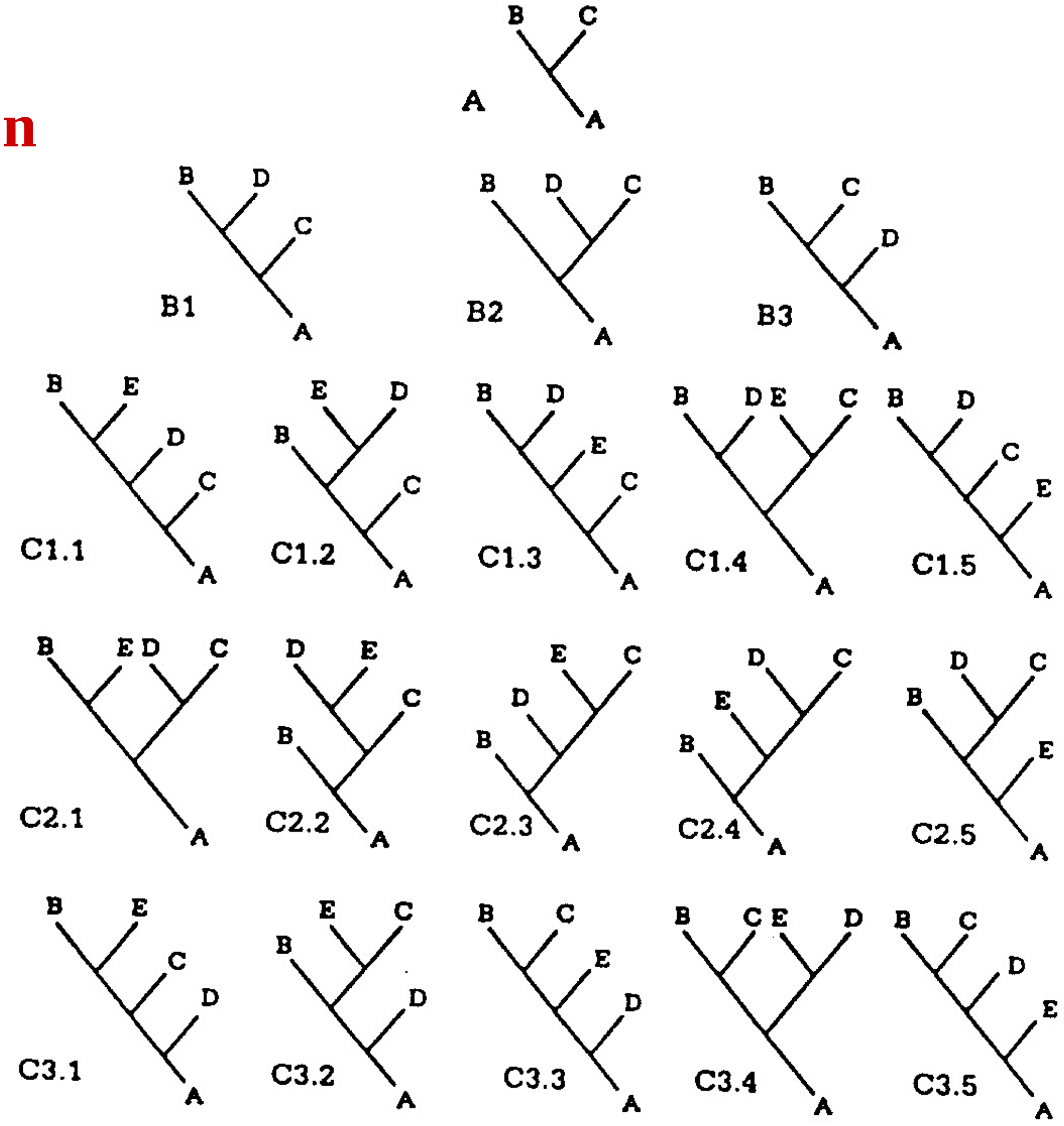
Quick-and-dirty methods, that do not grant that the shortest tree will be found

Improvement of the initial trees

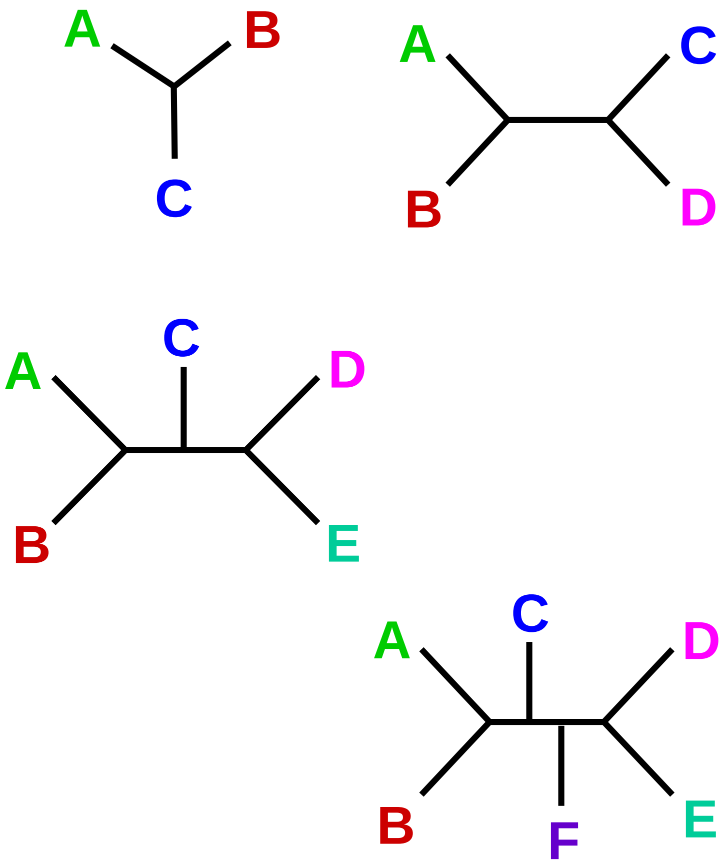
Branch swapping

Tree building methods

Exhaustive search, implicit enumeration

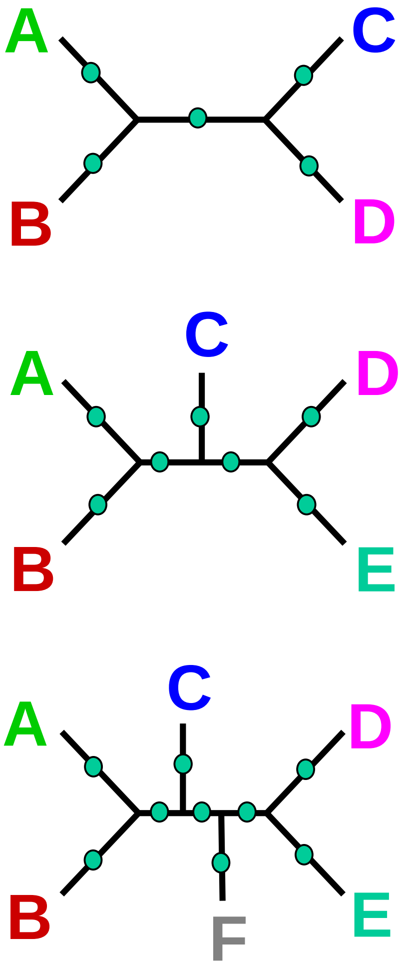


Exhaustive search makes sense up to ca. 11 taxa



# Taxa (N)	# Unrooted trees
3	1
4	3
5	15
6	105
7	945
8	10,935
9	135,135
10	2,027,025
.	.
.	.
.	.
.	.
30	3.58×10^{36}

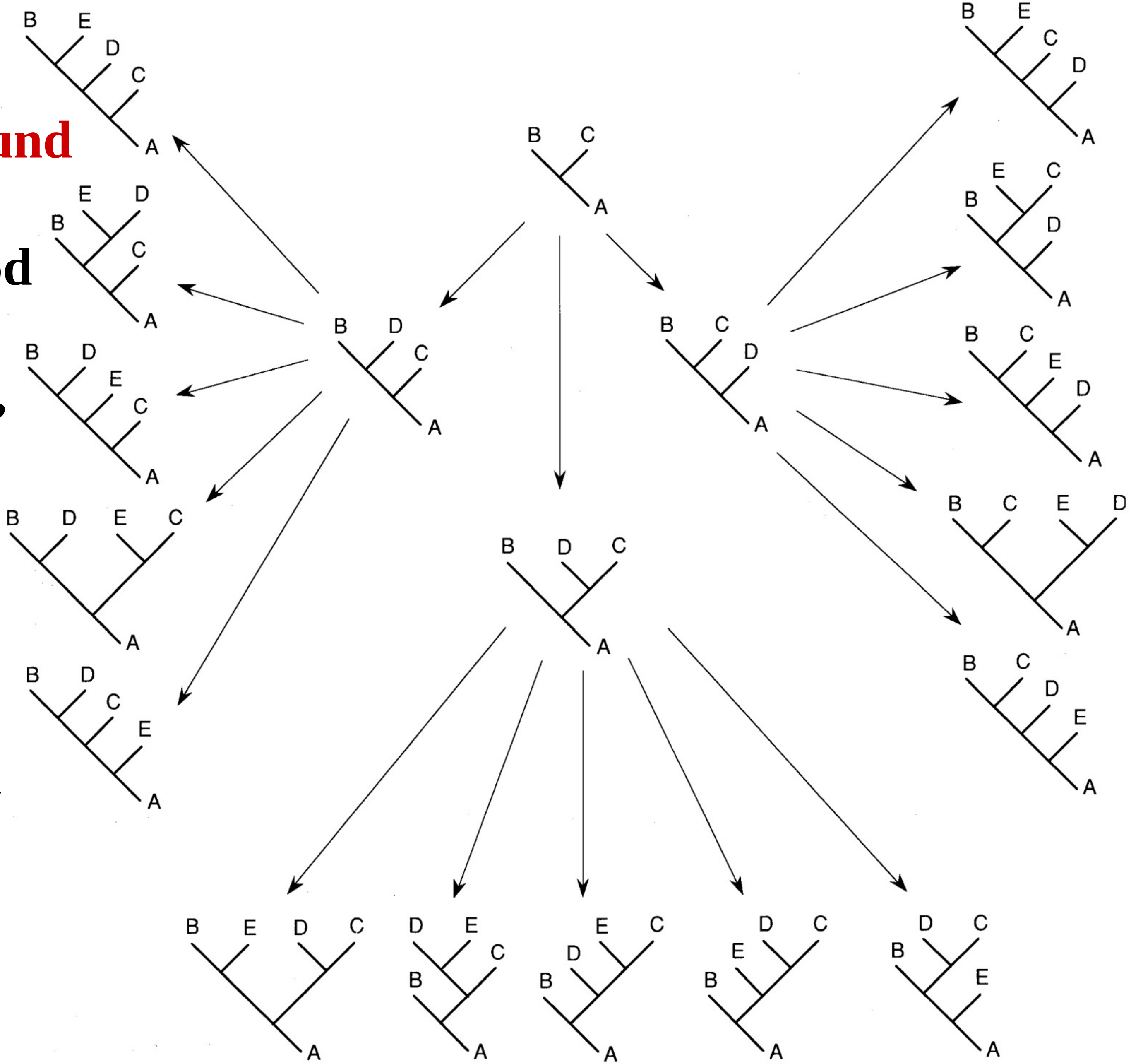
Each unrooted tree (theoretically) can be rooted anywhere along any of its branches



# Taxa	# Unrooted Trees	x ● # Roots	= # Rooted Trees
3	1	3	3
4	3	5	15
5	15	7	105
6	105	9	945
7	945	11	10,395
8	10,935	13	135,135
9	135,135	15	2,027,025
.	.	.	.
.	.	.	.
30	$\sim 3.58 \times 10^{36}$	57	$\sim 2.04 \times 10^{38}$
.	.	.	.
135			2.11×10^{267}

exceeds the number of particles in the known universe!!!

Tree building methods



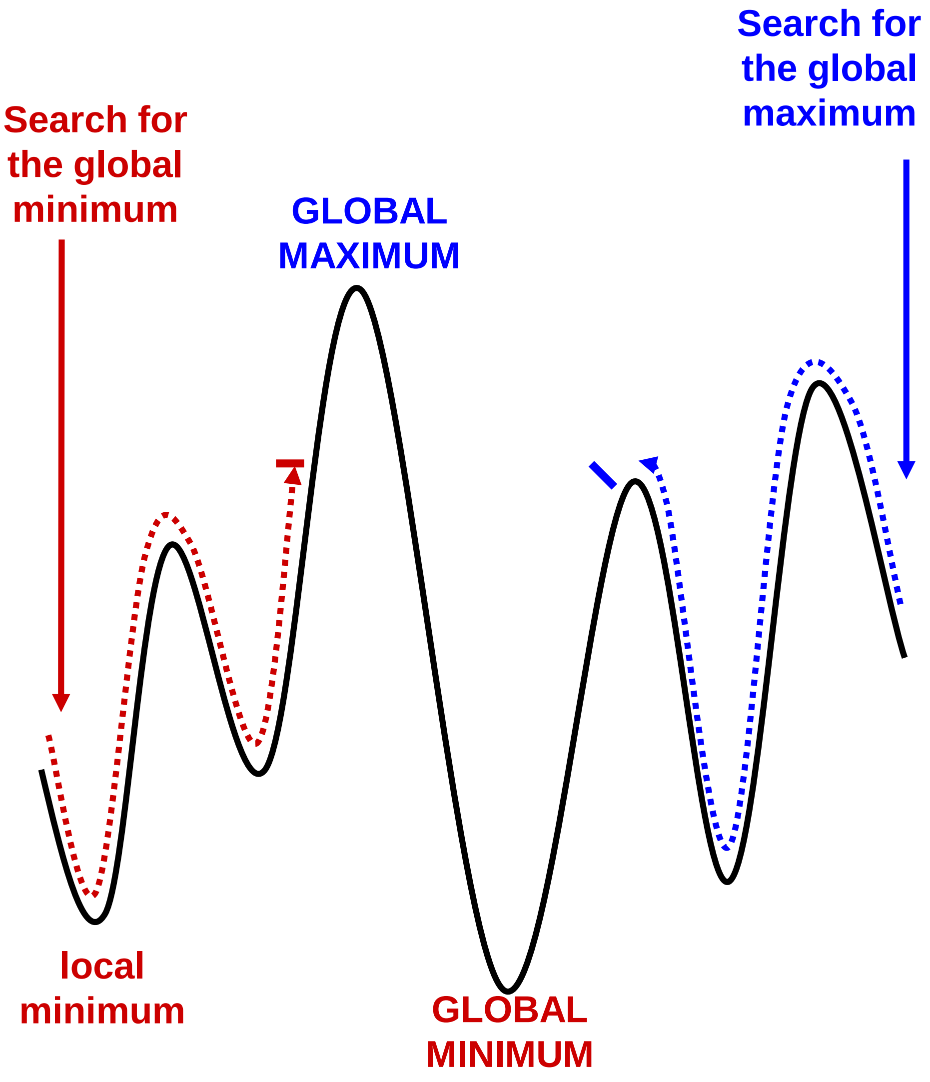
Branch-and-bound

Heuristic method is used to find a suboptimal tree, which serves as the first step (starting tree) for the analyses

Uo to ca 25 taxa

Heuristic algorithms

Quick-and-dirty methods, that do not grant that the shortest tree will be found



Search for the global minimum

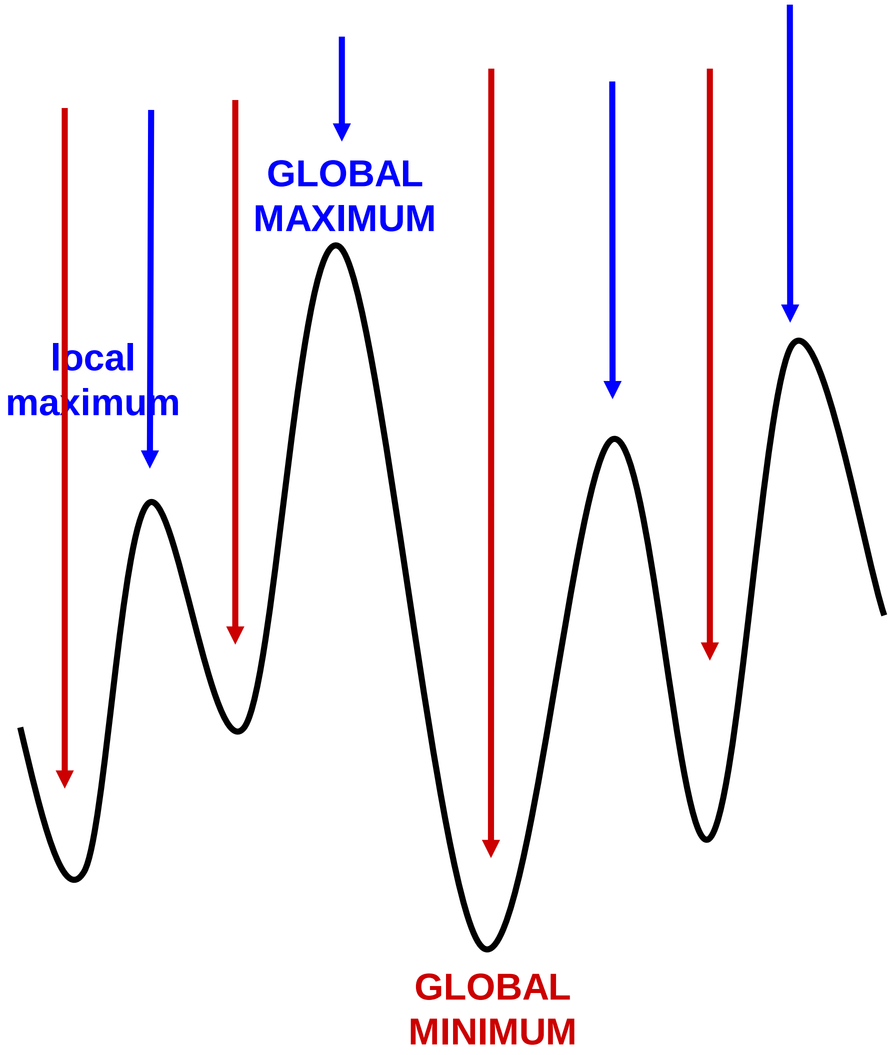
GLOBAL MAXIMUM

Search for the global maximum

local minimum

GLOBAL MINIMUM

Stepwise addition



local maximum

GLOBAL MAXIMUM

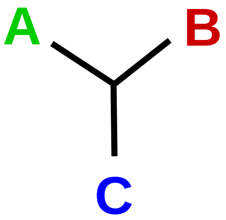
GLOBAL MINIMUM

Branch swapping

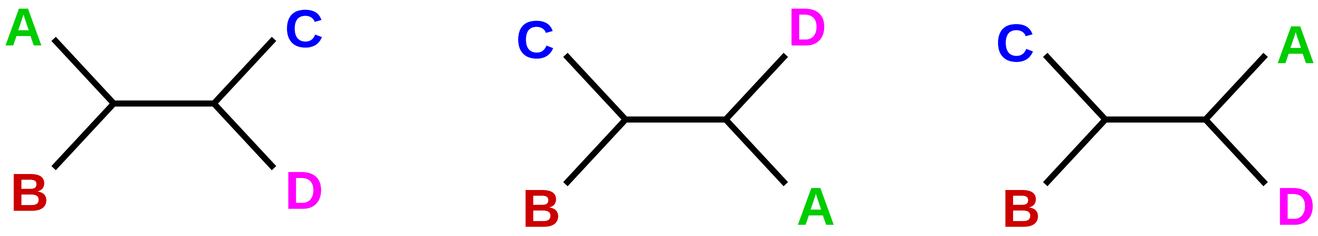
Heuristic algorithms

Stepwise addition

First, three taxa joined



Next, one of the unplaced taxa selected for next addition and placed along one of 3 branches.



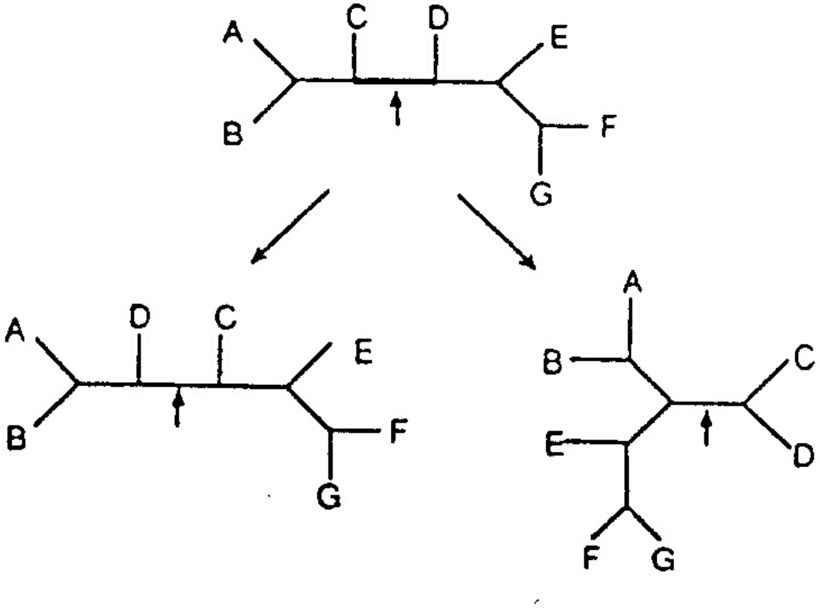
Each tree is evaluated by the optimality criterion and the best is retained for next round of addition.

Process continues until all objects are put on the tree

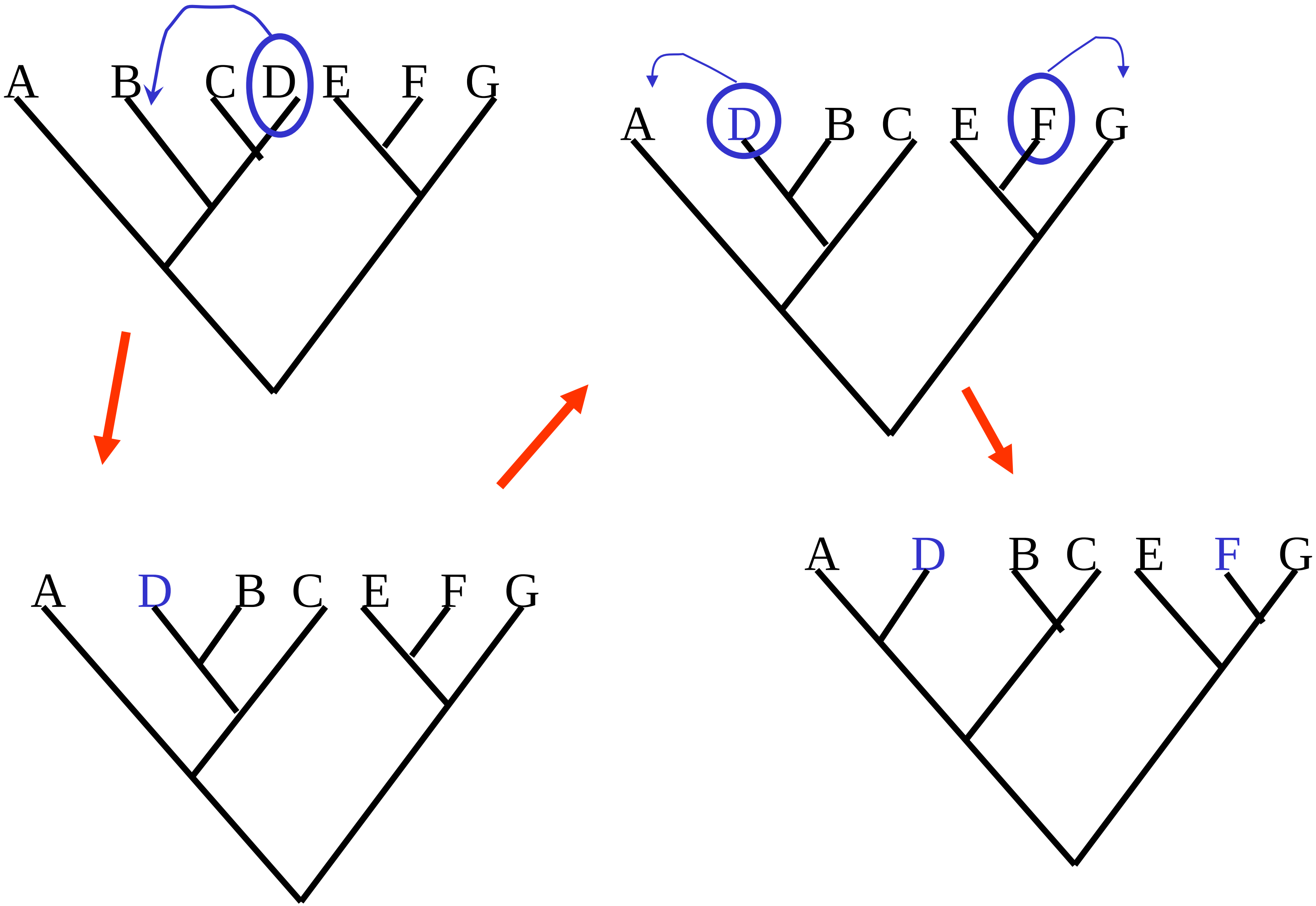
Heuristic search is repeated number of times, always with different imnitial three taxa – repeated replications (Nreps in PAUP)

Heuristic algorithm may start also with the Wagner tree, on which we apply branch swaping, while in each step we keep several trees that are further swaped.

Branch swapping

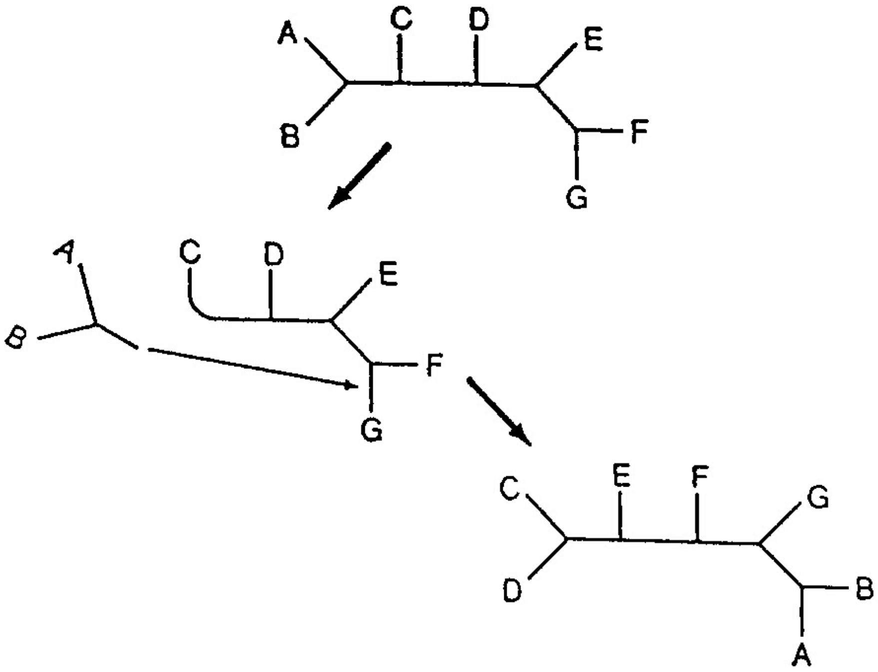


Nearest neighbor interchange (NNI)

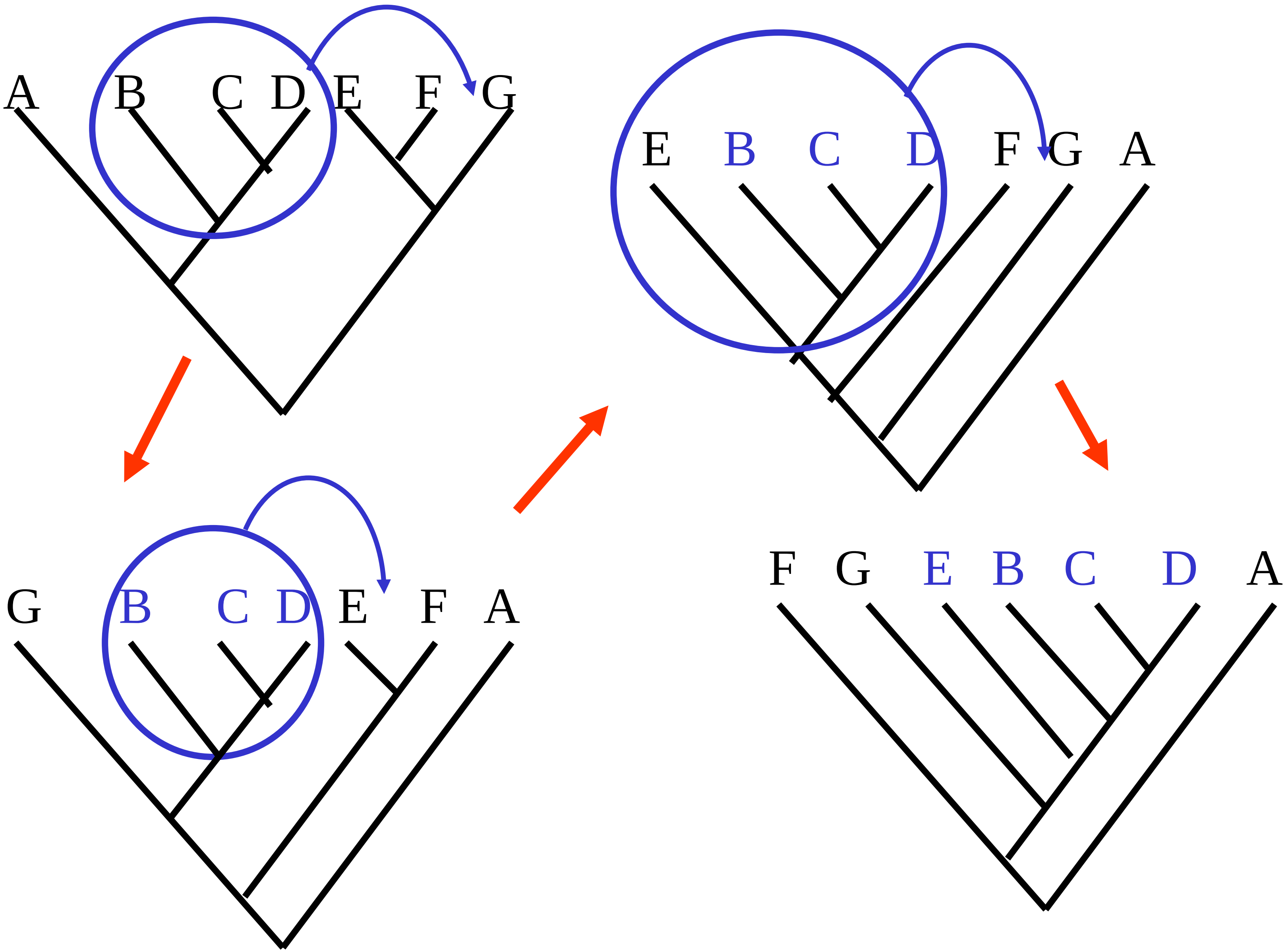


Nearest neighbor interchange (NNI)

Branch swapping

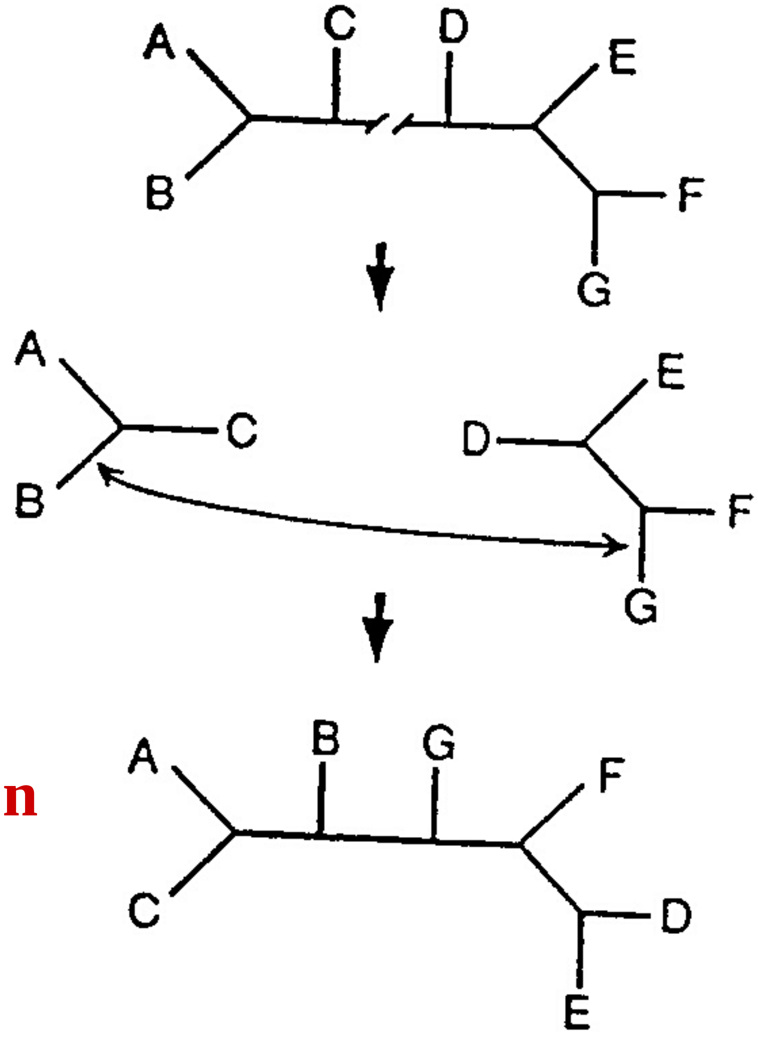


Subtree pruning and regrafting (SPR)

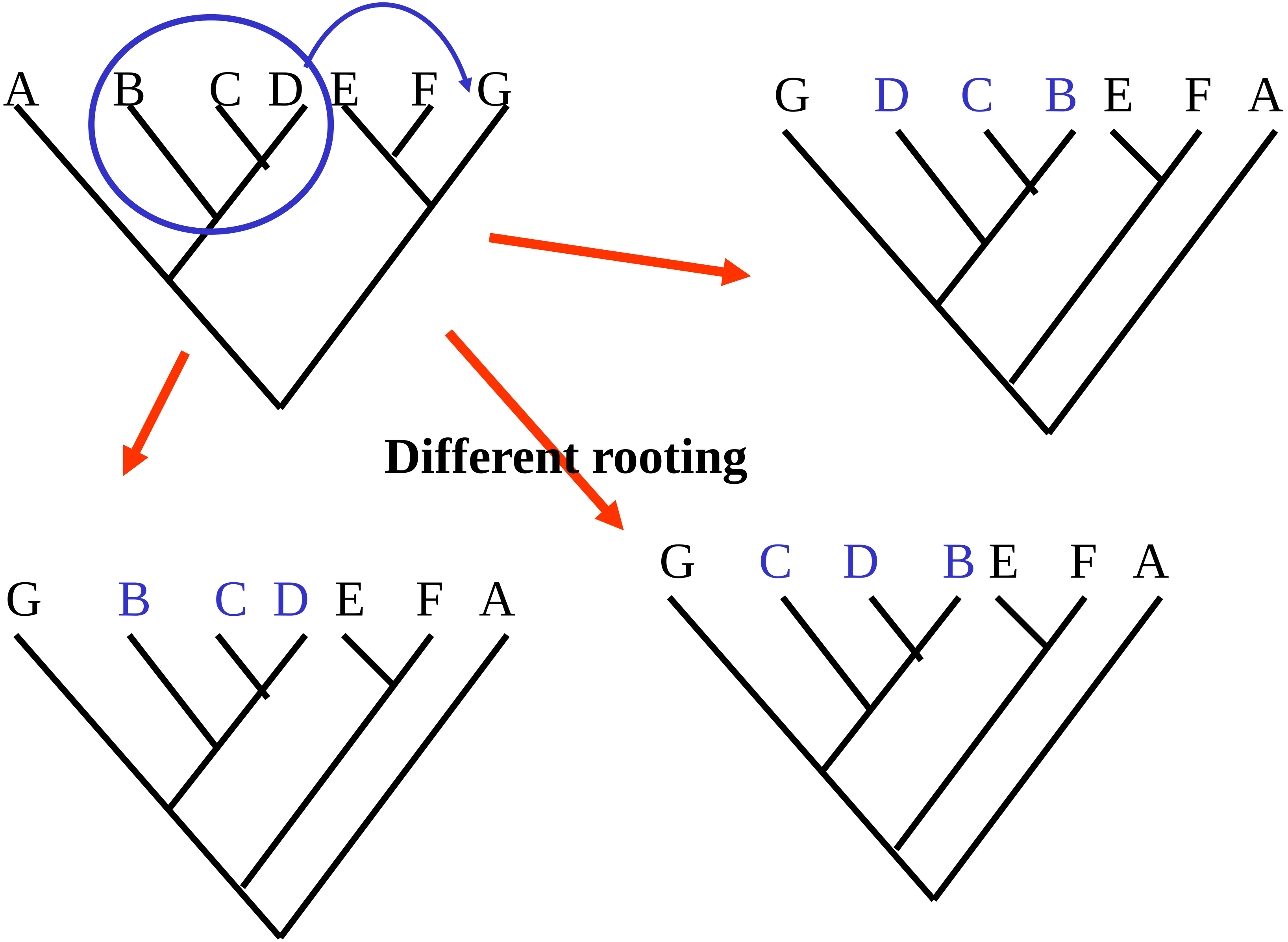


Subtree pruning and regrafting (SPR)

Branch swapping



Tree bisection and reconnection (TBR)



Different rooting

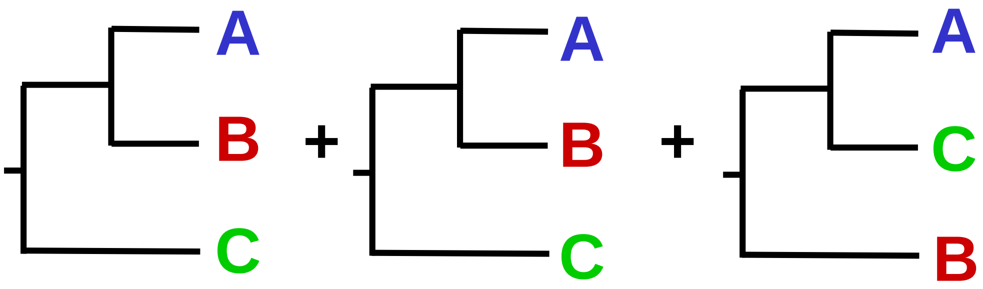
Tree bisection and reconnection (TBR)

TBR explores more trees than SPR

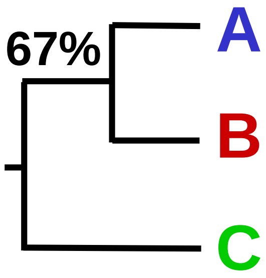
TBR is more effective in searching for a short three, but search takes more time

TBR is not the most effective method if we swap the tree, which is rather far from the shortest one – potentially each SPR swap is able to find short tree in such case

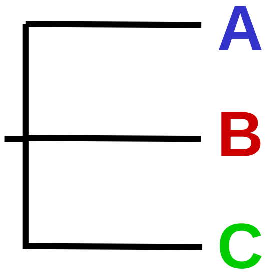
Consensus trees



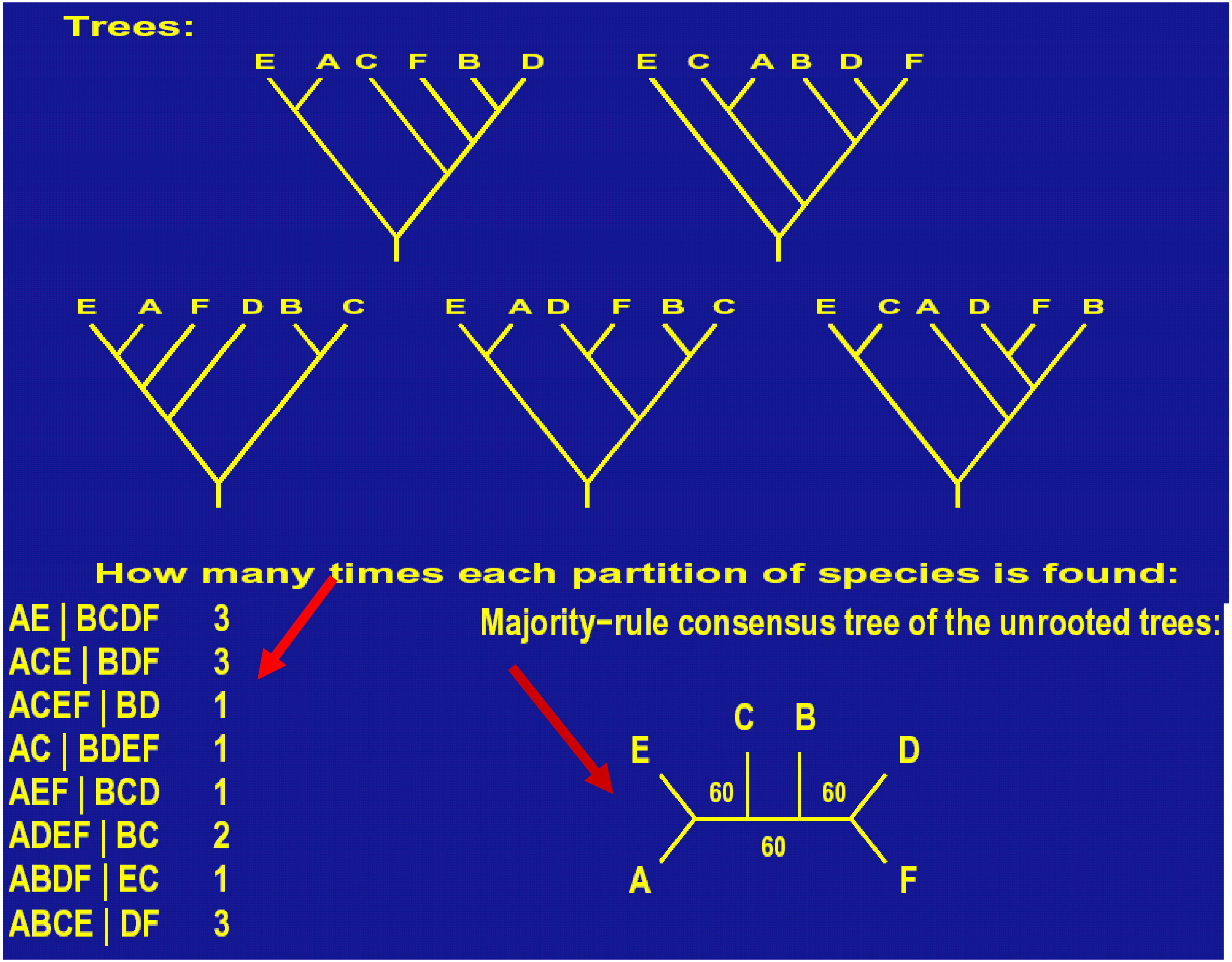
Majority-rule



Strict consensus trees



Majority-rule



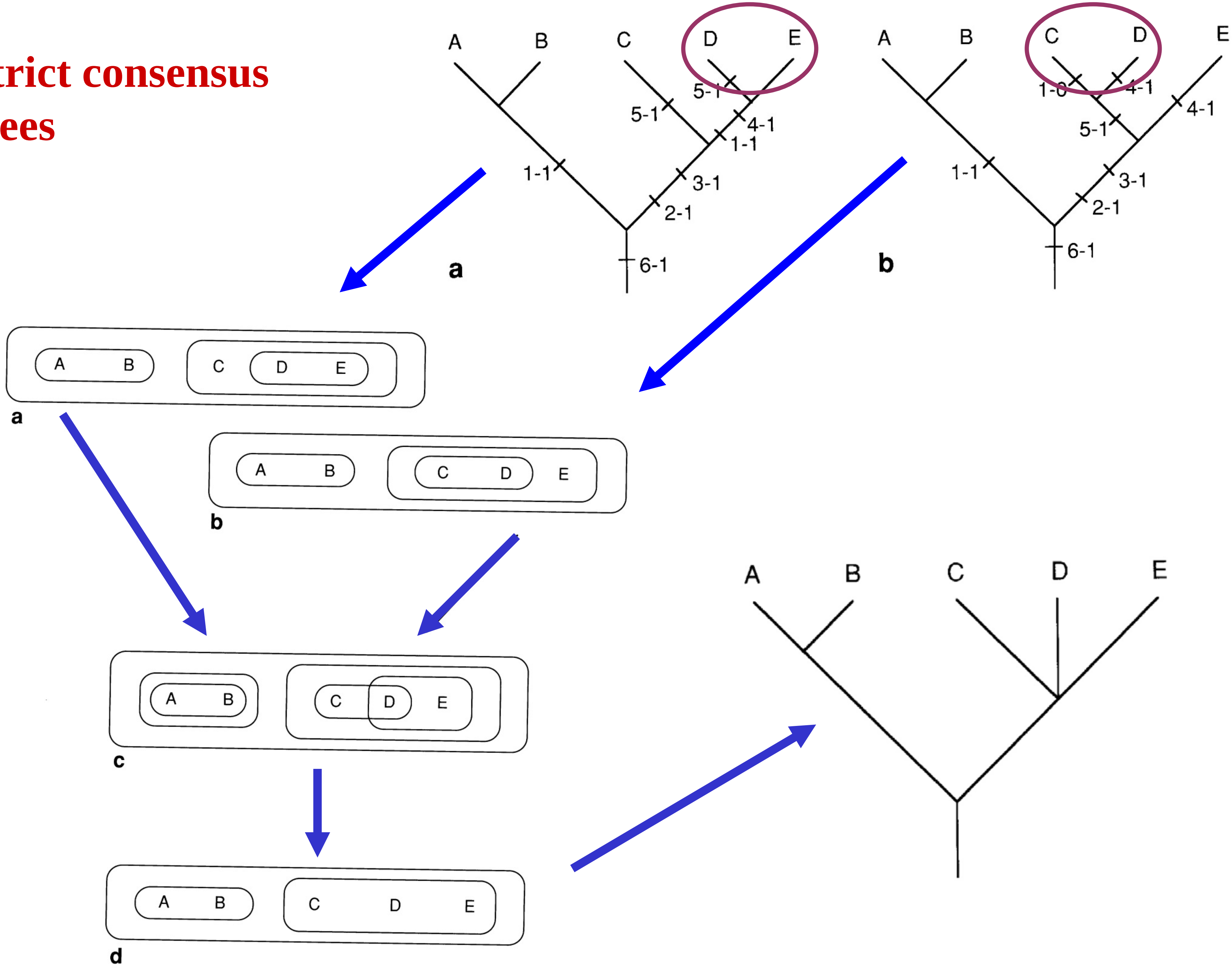
Trees:

How many times each partition of species is found:

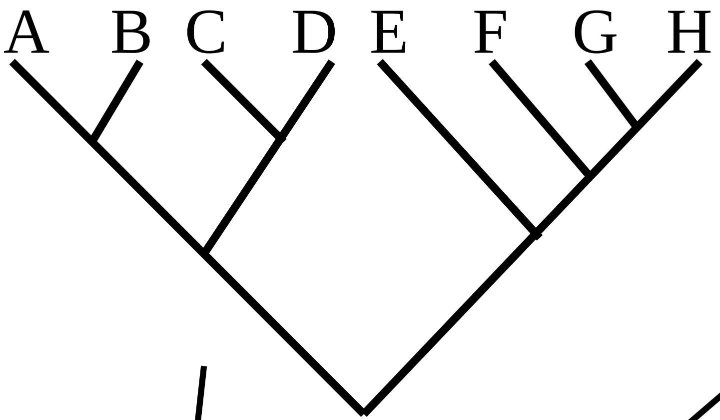
AE | BCDF	3
ACE | BDF	3
ACEF | BD	1
AC | BDEF	1
AEF | BCD	1
ADEF | BC	2
ABDF | EC	1
ABCE | DF	3

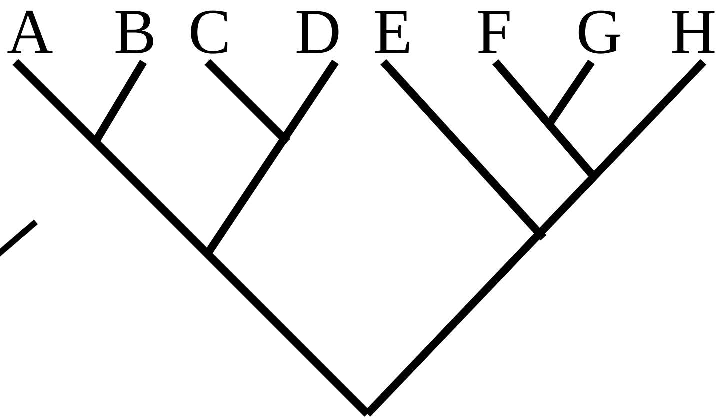
Majority-rule consensus tree of the unrooted trees:

Strict consensus trees



Strict consensus trees

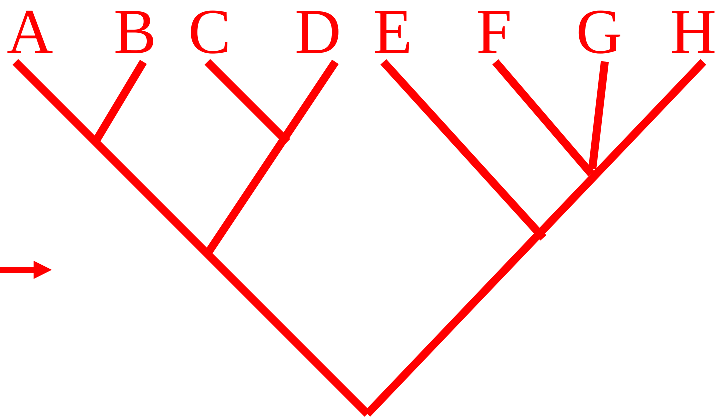




AB
CD
ABCD
GH
FGH
EFGH

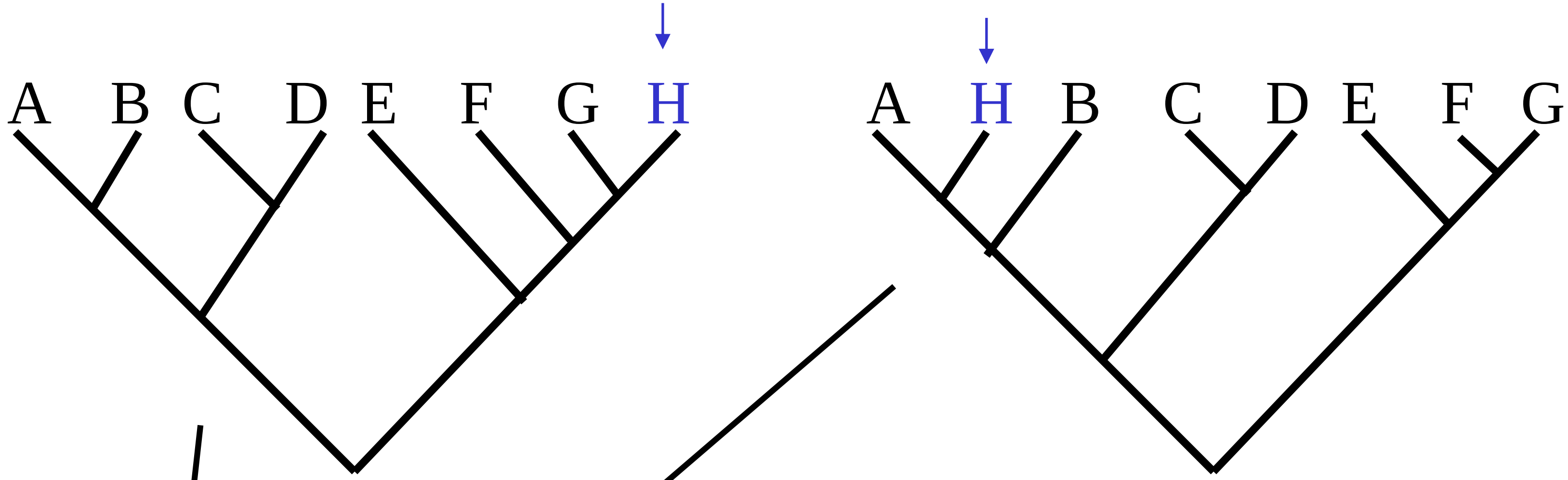
AB
CD
ABCD
FG
FGH
EFGH

components



Strict consensus tree

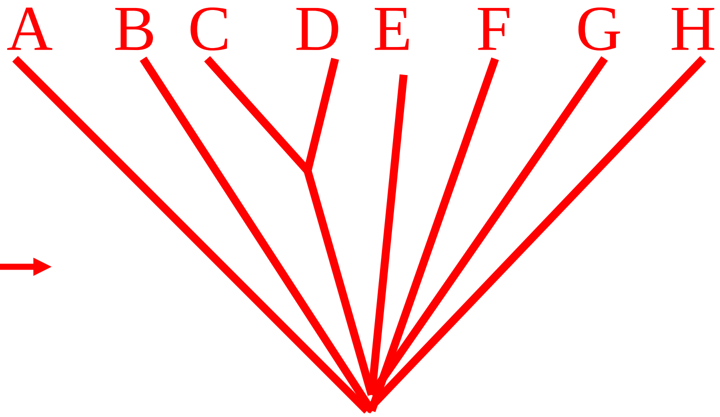
Uncertain position of a single individual may cause colaps of the tree (even strong signal in data can be overturn by one problematic individual)



AB
CD
ABCD
GH
FGH
EFGH

AH
ABH
CD
ABCDH
FG
EFGH

Components



Strict consensus tree

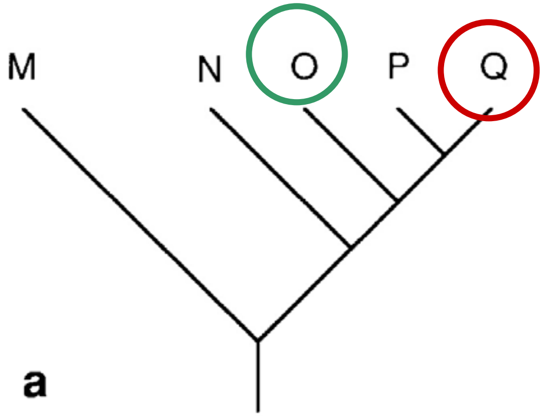
Polytomy is a consequence of:

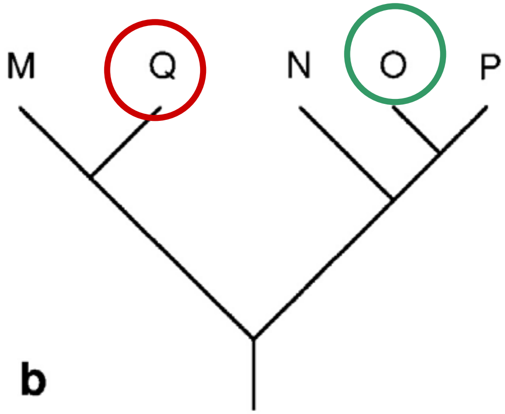
1. Insufficient data – we do not have enough variation to get reliable tree

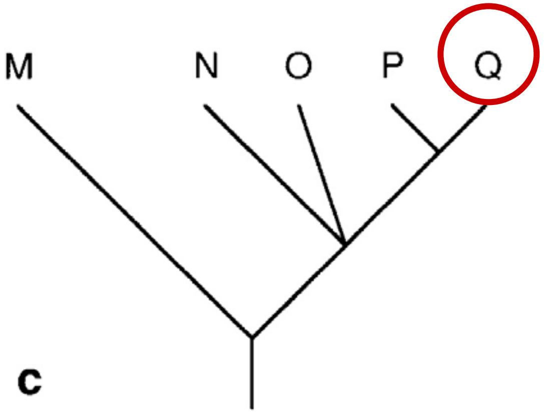
2. Conflict in data – characters equally support position of the taxon on more than one place on the tree (=> missing data)

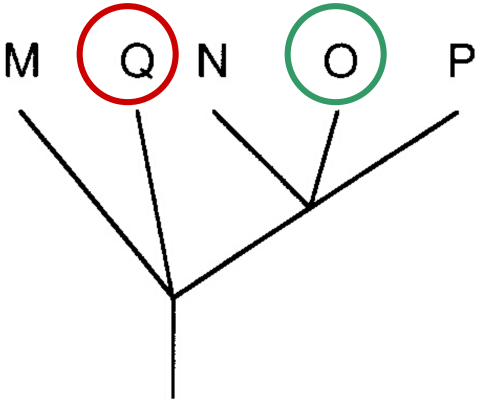
Polytomy should not be interpreted as rapid radiation – there is an uncertainty in the data!!!

Adams consensus trees

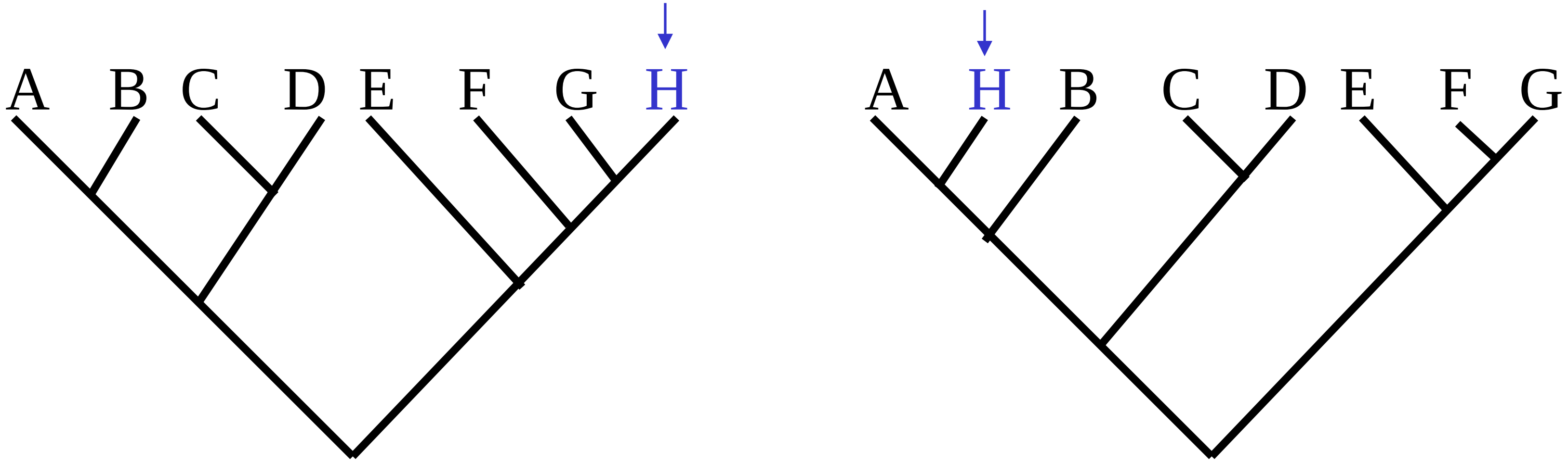


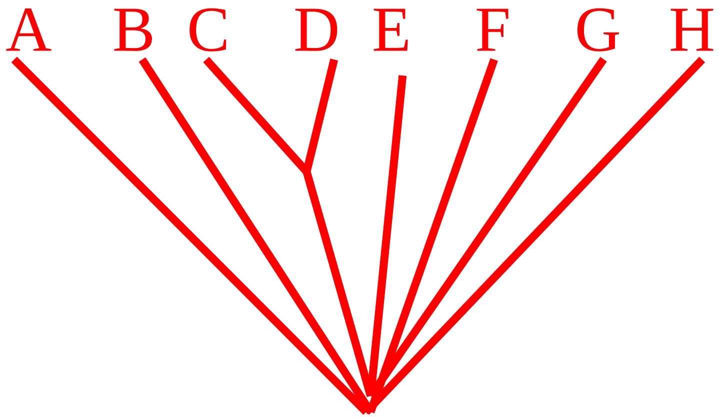




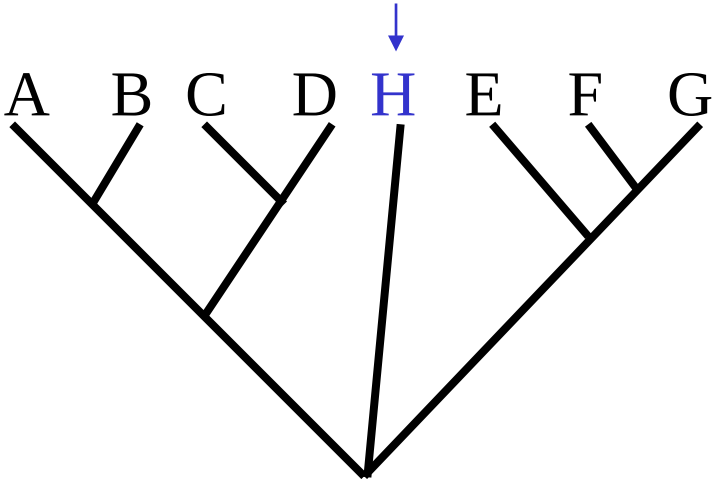


Taxa causing conflicts are moved to the node, where they do not cause conflict with any tree





Strict consensus tree



Adams consensus tree

Bootstrap

- Uses random sampling with replacement (we create replicated data matrices)
- We analyze each replicated data set
- Looking for the same branches that appear on the tree

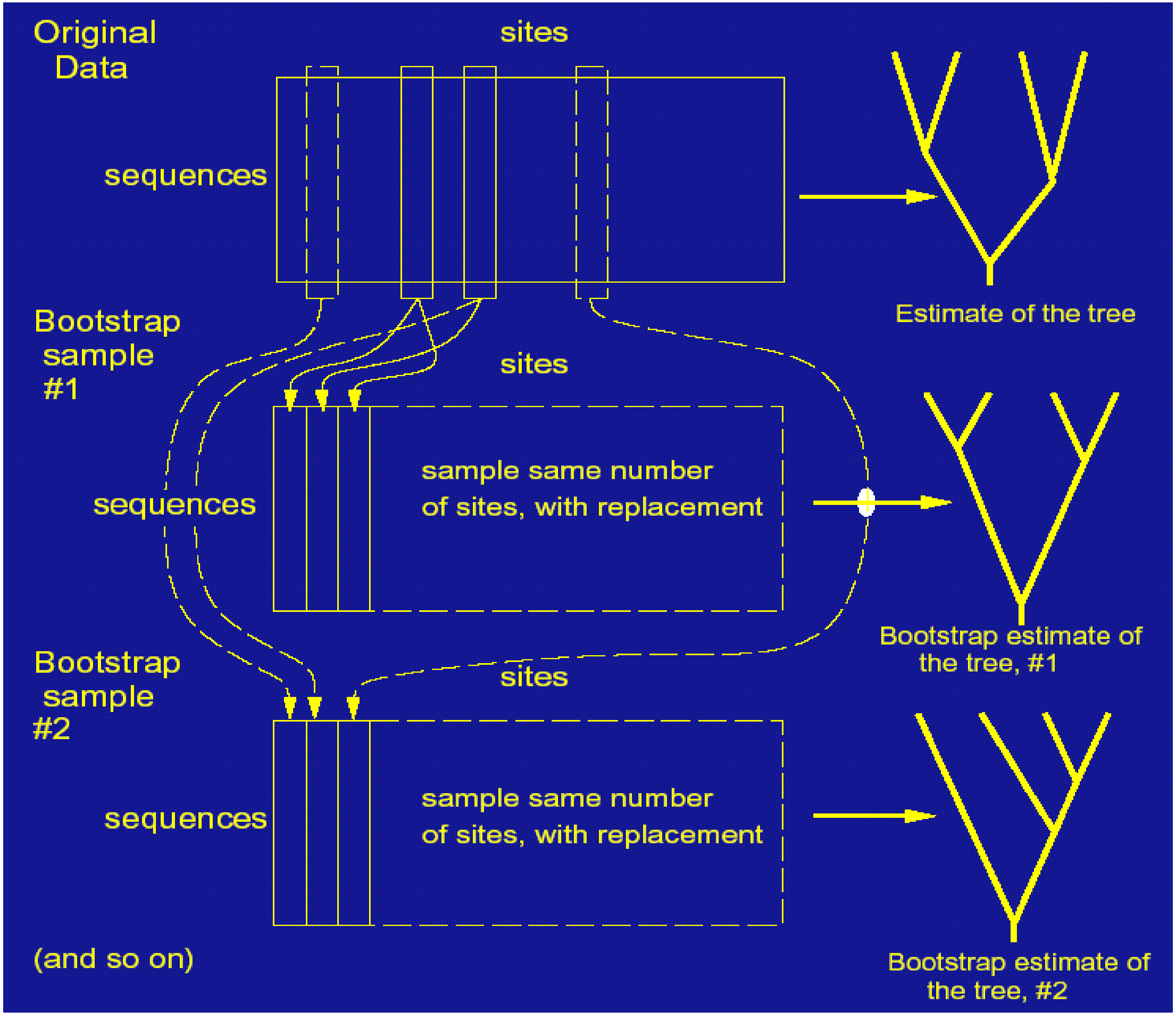
original data matrix

	znaky							
taxóny	1	2	3	4	5	6	7	8
A	R	R	Y	Y	Y	Y	Y	Y
B	R	R	Y	Y	Y	Y	Y	Y
C	Y	Y	Y	Y	Y	R	R	R
D	Y	Y	R	R	R	R	R	R
Outgp	R	R	R	R	R	R	R	R

replicated data matrix

	znaky							
taxóny	1	2	2	5	5	6	6	8
A	R	R	R	Y	Y	Y	Y	Y
B	R	R	R	Y	Y	Y	Y	Y
C	Y	Y	Y	Y	Y	R	R	R
D	Y	Y	Y	R	R	R	R	R
Outgp	R	R	R	R	R	R	R	R

Bootstrap


Original
Data
sites
sequences
Estimate of the tree
Bootstrap
sample
#1
sites
sequences
sample same number
of sites, with replacement
Bootstrap estimate of
the tree, #1
Bootstrap
sample
#2
sites
sequences
sample same number
of sites, with replacement
Bootstrap estimate of
the tree, #2
(and so on)

Jack-knife

- Jack-knife is similar to the bootstrap, it differs only in the strategy of selection of characters
- Randomly selected characters are deleted from the matrix (e.g., 50%)
- Jack-knife seldom occurs in statistical packages and is also seldom used
- Jack-knife and bootstrap have tendency to produce similar results

Bremer support = decay analysis

Bremer support can be expressed as a number of further steps on the tree that result in the collapse of the branch into polytomy in the strict consensus tree.

Branches on the tree that collapse during the building of the strict consensus tree from all most parsimonious (shortest) trees possess value of the Bremmer support = 0, such support is usually not presented (neither branched with such support).

When we add to all most parsimonious (shortest) trees all trees that are one step longer, some branches will collapse, they possess value of Bremmer support = 1, etc.

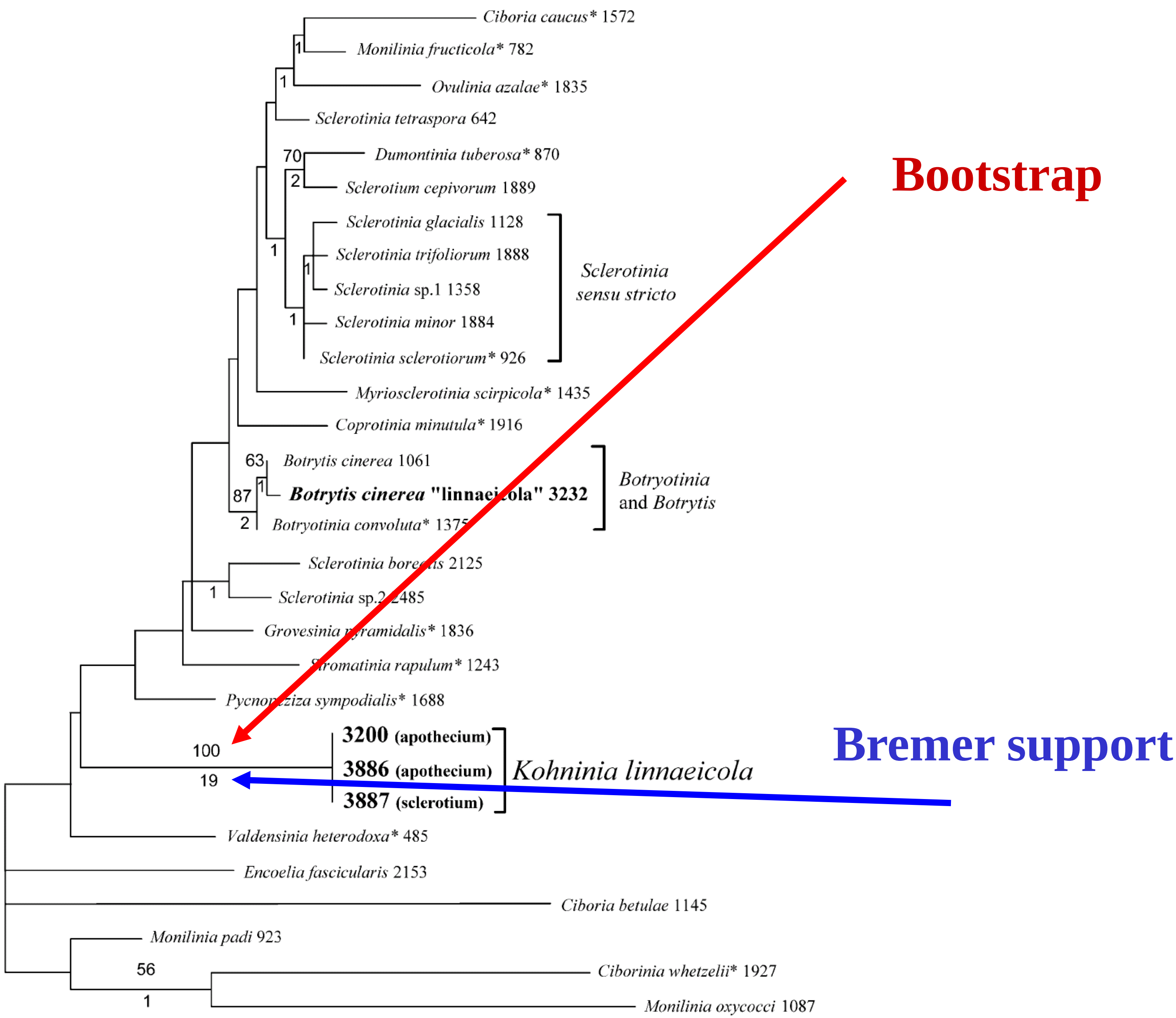
Bremer support = decay analysis

There are two possible strategies of the calculation of the Bremer support:

(1) Calculate all cladograms that are certain amount of steps longer than most parsimonious trees and search at which length of the tree given branch collapses (disappears).

(2) Calculate all possible trees that do not contain given branch and measure their length (metóda „reverse constraints“)

ATTENTION: computing of the Bremer support is time consuming, if interrupted earlier acquired values of Bremer support are too high



Bootstrap

Bremer support

Rules of Phylogenetic Classifications

1. Only monophyletic groups will be formally classified

2. All classifications will be logically consistent with the phylogenetic hypothesis accepted by the investigator

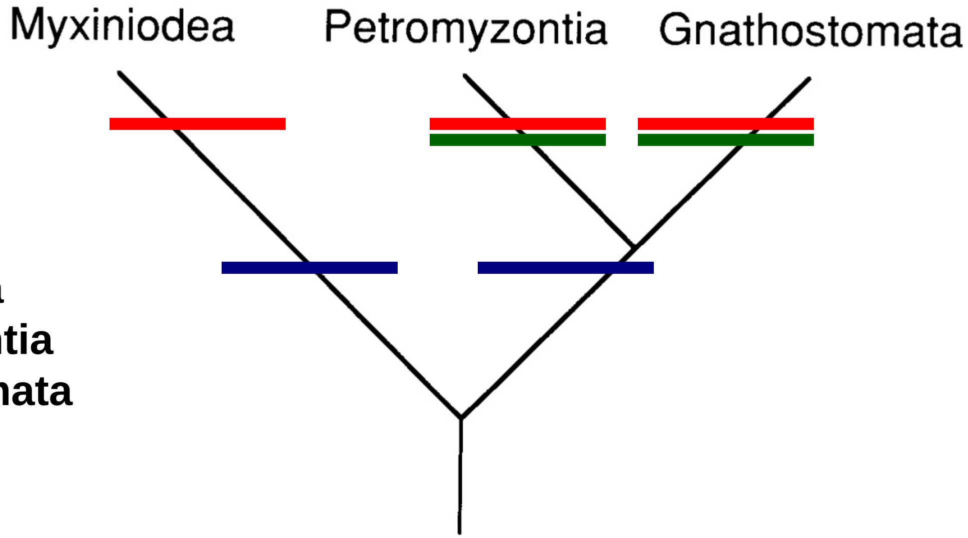
3. Regardless of the conventions used, each classification must be capable of expressing the sister group relationships among the taxa classified

Conventions

1. The Linnaean system of ranks will be used

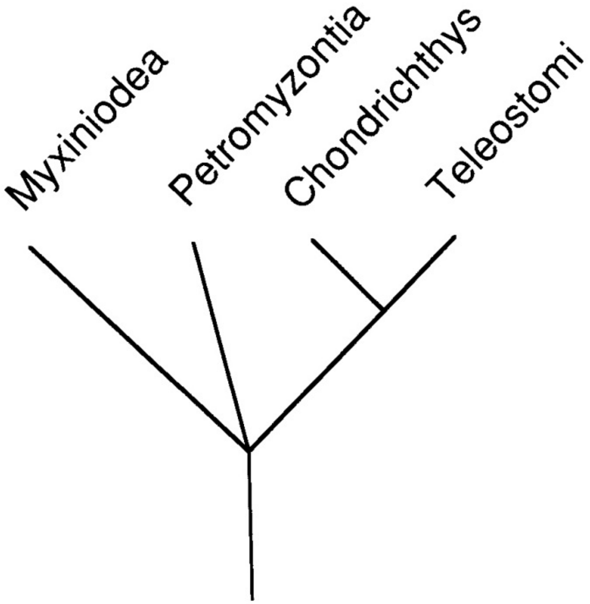
2. Minimum taxonomic decisions will be made to construct a classification or to modify existing classifications.

3. Taxa forming an asymmetrical part of a phylogenetic tree may be placed at the same rank and sequenced in their order of branching; When such a list is encountered, the sequence of the list denotes the sequence of the branching.



Subphyllum Vertebrata
    Infraphyllum Myxinioidea
    Infraphyllum Petromyzontia
    Infraphyllum Gnathostomata

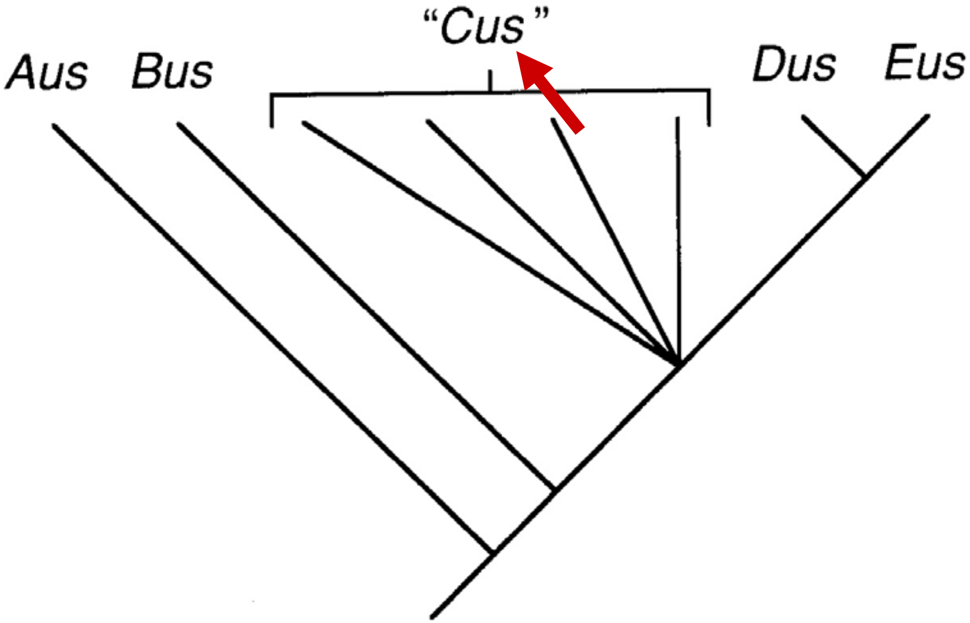
4. Taxa whose relationships are polytomous will be placed *sedis mutabilis* at the same rank



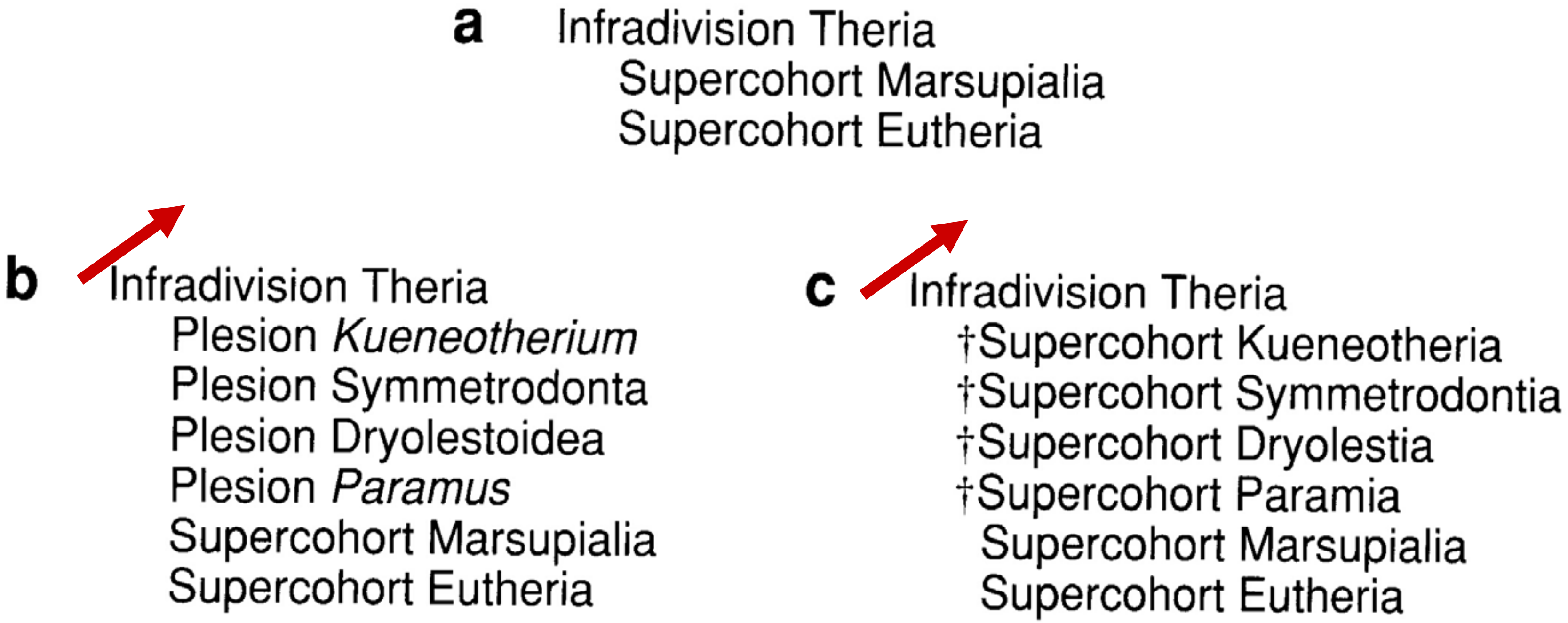
Subphyllum Vertebrata
- **Infraphyllum Myxinioidea (*sedis mutabilis*)**
- **Infraphyllum Petromyzontia (*sedis mutabilis*)**
- **Infraphyllum Gnathostomata (*sedis mutabilis*)**
  - **Superclass Chondrichthys**
  - **Superclass Teleostomi**

5. Monophyletic taxa of uncertain relationships will be placed *incertae sedis* at a level in the hierarchy where their relationships are known with some certainty

6. A group whose qualities are not known may be included in a phylogenetic classification if it is treated as *incertae sedis* and its name is put in shutter quotes (quotation marks)



7. Fossil taxa will be treated differently than Recent taxa. Fossil taxa will always be sequenced with their Recent relatives following Convention 3. If they are ranked, their status as fossils will be denoted by placing a dagger or cross symbol before the rank (Nelson, 1972). Alternatively, they may be given the neutral rank of “plesion”.As natural taxa, monophyletic fossil taxa may stand *incertae sedis* or *sedis mutabilis*, just as any Recent natural taxon.



a

Infradivision Theria
  Supercohort Marsupialia
  Supercohort Eutheria

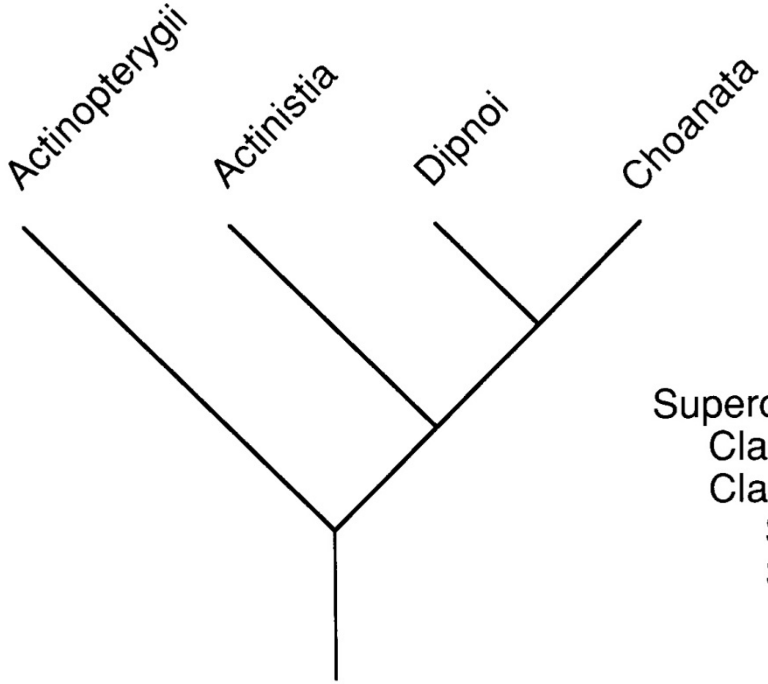
b

Infradivision Theria
  Plesion *Kueneotherium*
  Plesion Symmetrodonta
  Plesion Dryolestoidea
  Plesion *Paramus*
  Supercohort Marsupialia
  Supercohort Eutheria

c

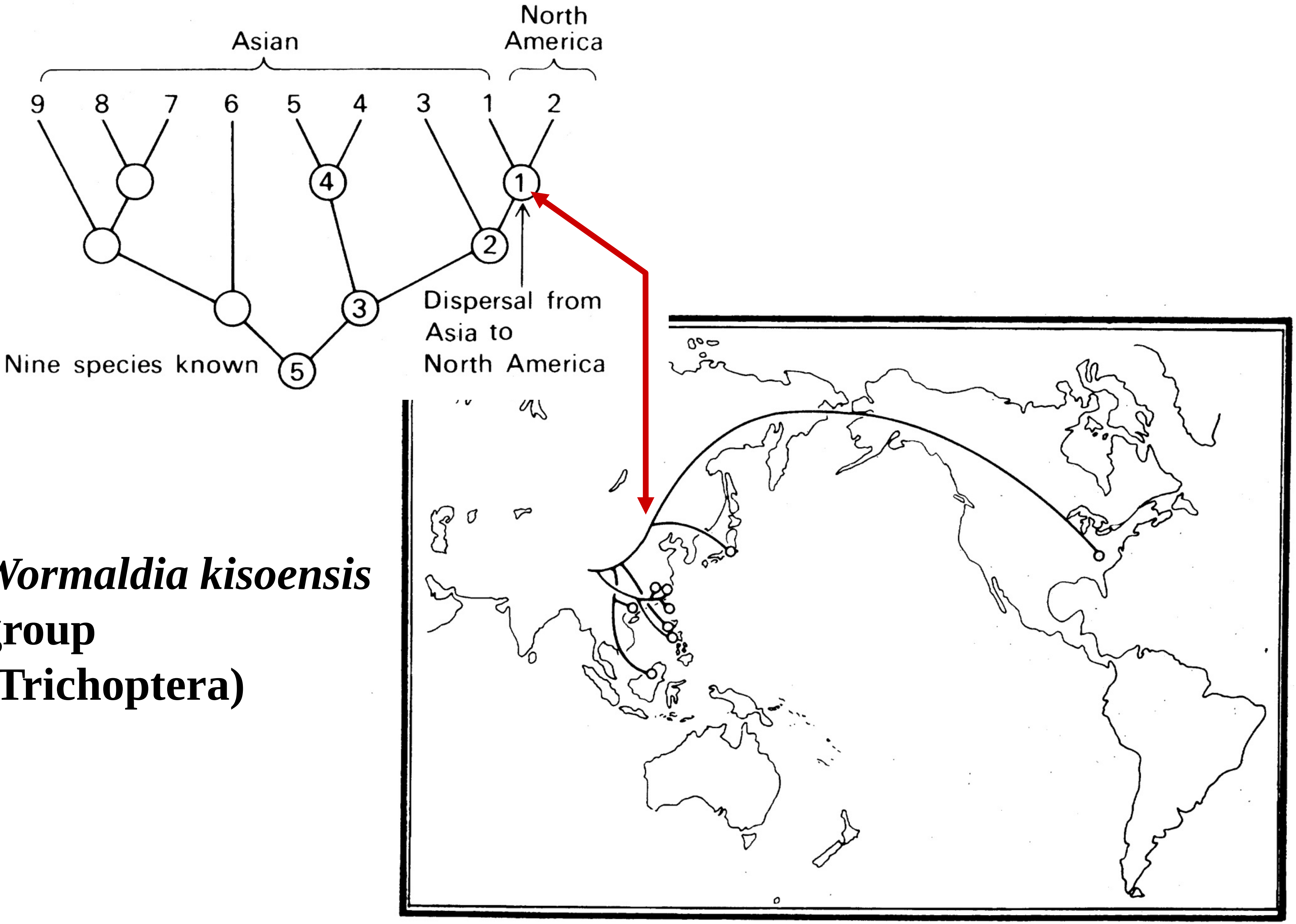
Infradivision Theria
  †Supercohort Kueneotheria
  †Supercohort Symmetrodontia
  †Supercohort Dryolestia
  †Supercohort Paramia
  Supercohort Marsupialia
  Supercohort Eutheria

8. Stem species (ancestral species) are placed in classifications in parentheses beside the names of taxa they gave rise to or taxa containing their descendants, as appropriate.



- Superclass Teleostomi
  - Class Actinopterygii
  - Class Sarcopterygii (*Sarcopterygius primus*)
    - Subclass Actinistia
    - Subclass Dipnoiformes
      - Infraclass Dipnoi
      - Infraclass Choanata

Cladistic biogeography



Wormaldia kisoensis **group (Trichoptera)**