

20.xi.

fossils & missing information

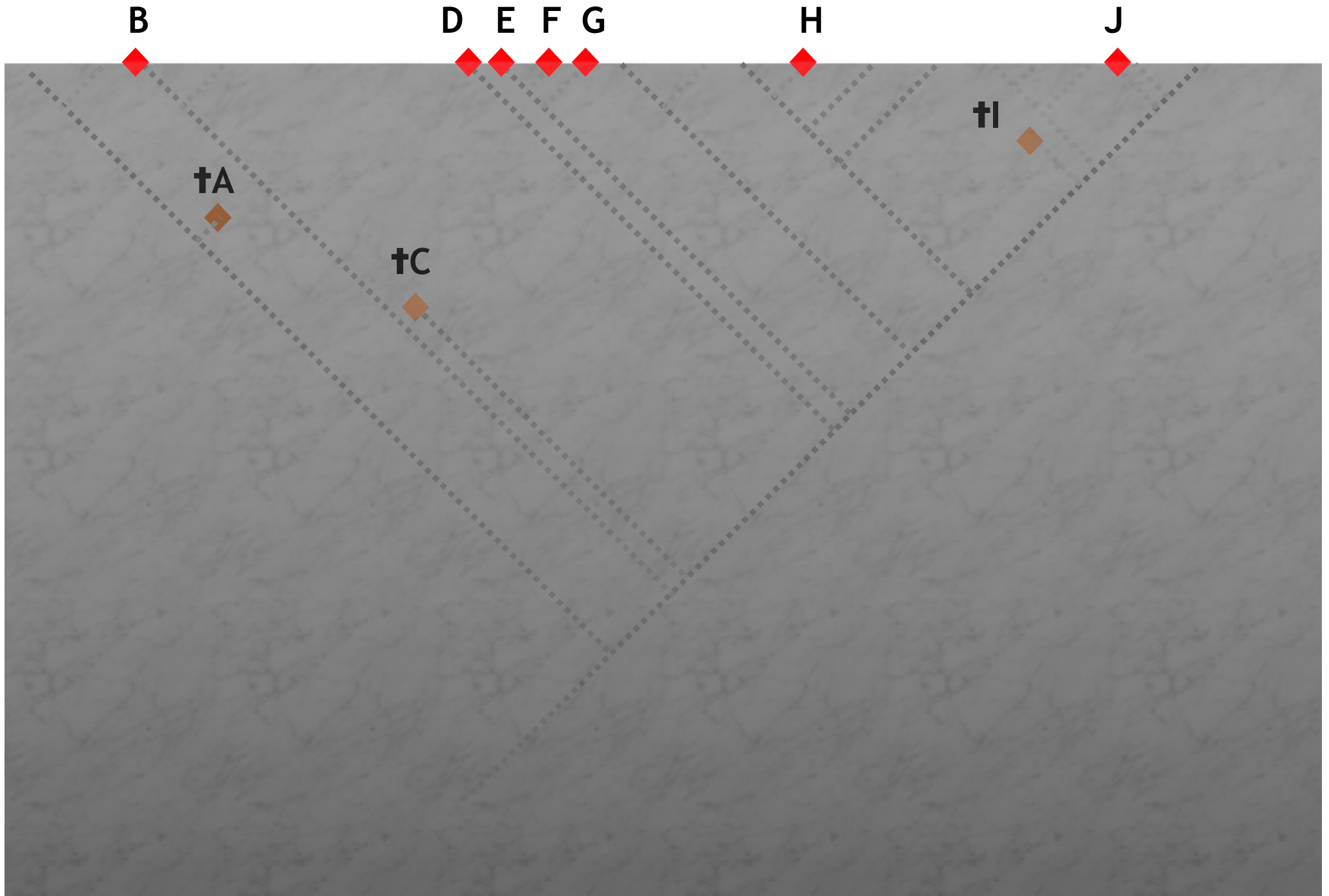
SMALL amount of information

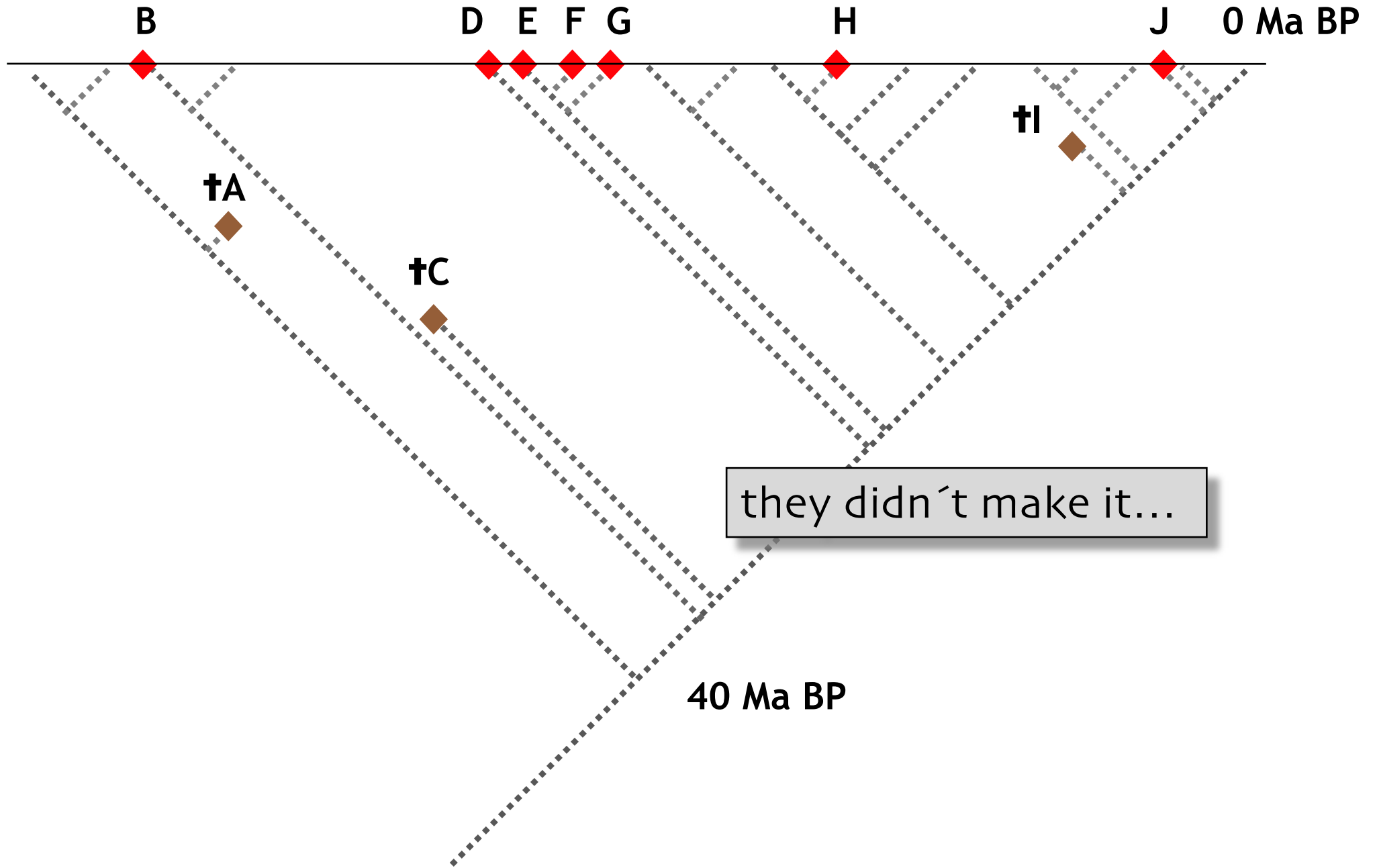
FOSSILS IN CLADISTIC ANALYSIS

EXTINCT organisms: 99,999..... % of global biodiversity through the history of life

NOVACEK, M. J., AND Q. D. WHEELER. 1992. Extinct taxa: accounting for 99.999...% of the earth's biota. Pages 1–16 *in* Extinction and phylogeny (M. J. Novacek and Q. D. Wheeler, eds.). Columbia Univ. Press, NY.

BASIC difference between extinct (fossils) & extant organisms?





FOSSILS IN CLADISTIC ANALYSIS

EXTINCT organisms: 99,999..... % of global biodiversity through the history of life

BASIC difference between extinct (fossils) & extant organisms?

should be treated in analysis AS ANY OTHER ORGANISMS

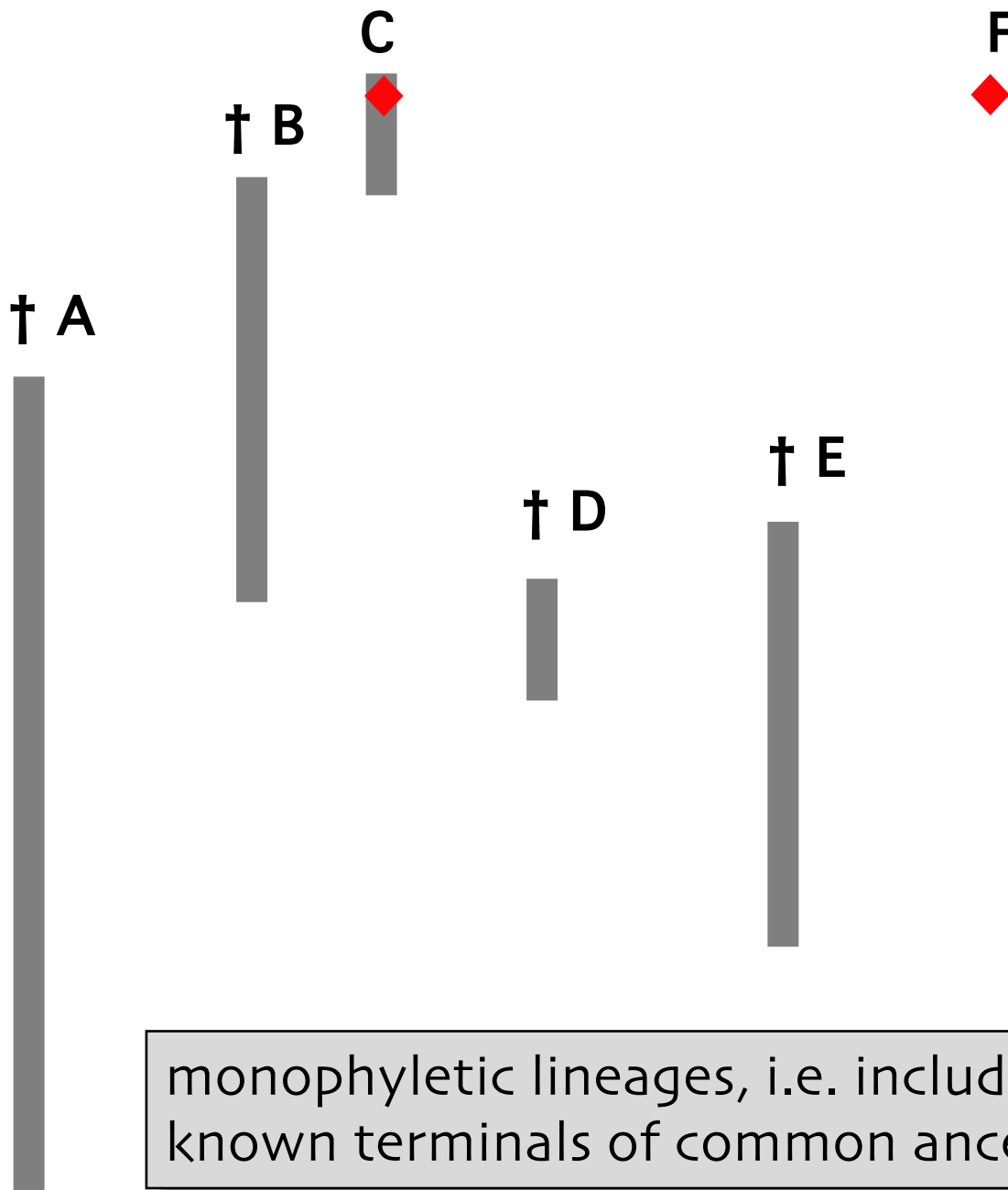
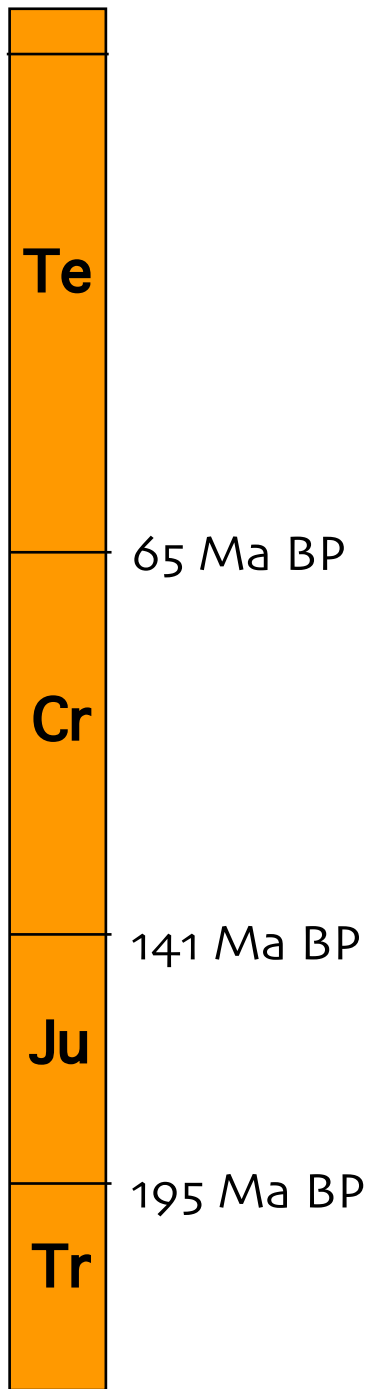
fossils are **NOT** necessary for hypotheses about phylogeny

but they can be useful in many ways...

and also problematic...

FOSSILS IN CLADISTIC ANALYSIS

minimum age for clades



hypothesis of phylogeny
of lineages

Te

† A

† B

C

sister group rule

CONDITIONAL on
the particular
hypothesis about
phylogeny

F

65 Ma BP

† D

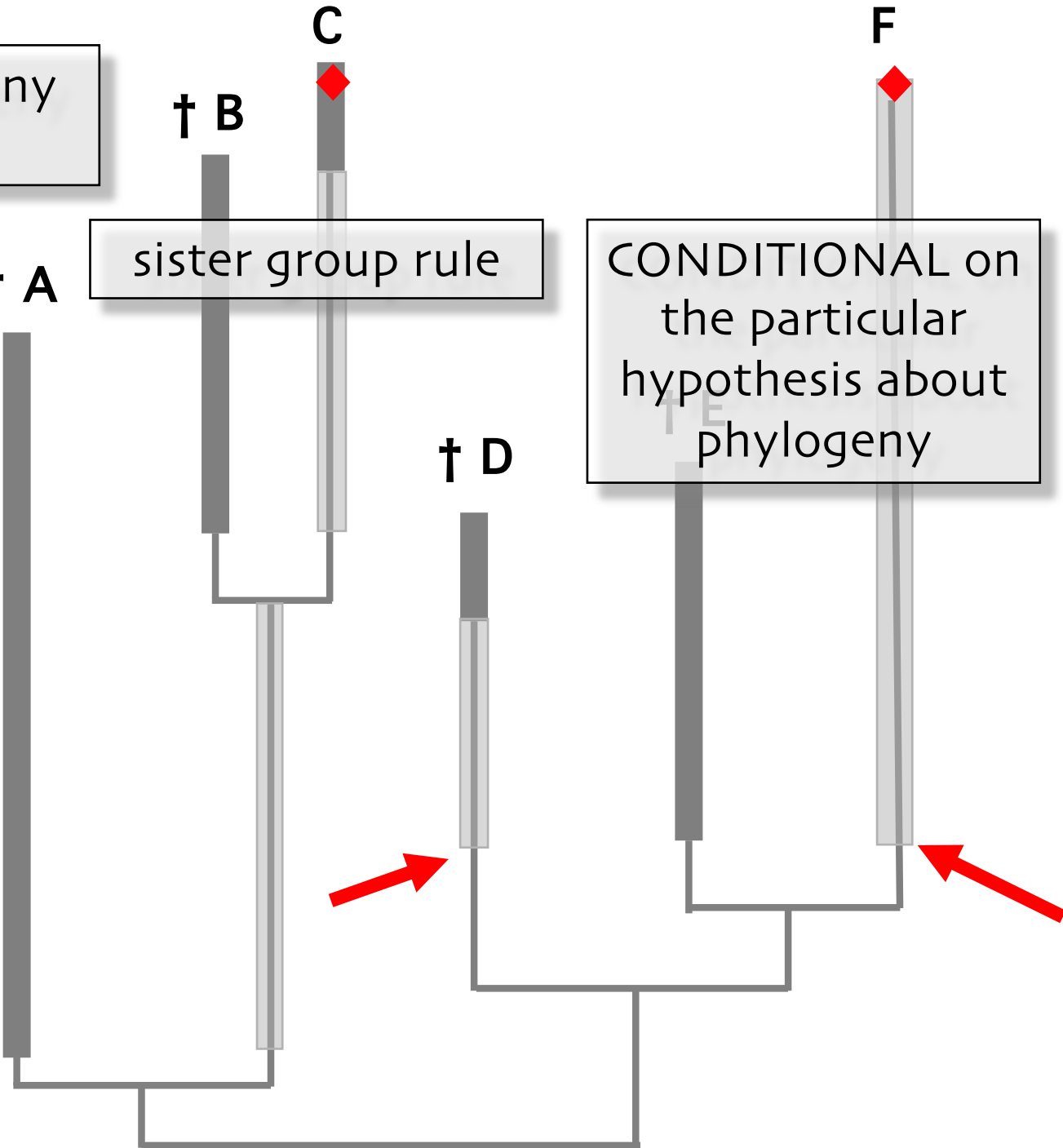
Cr

141 Ma BP

Ju

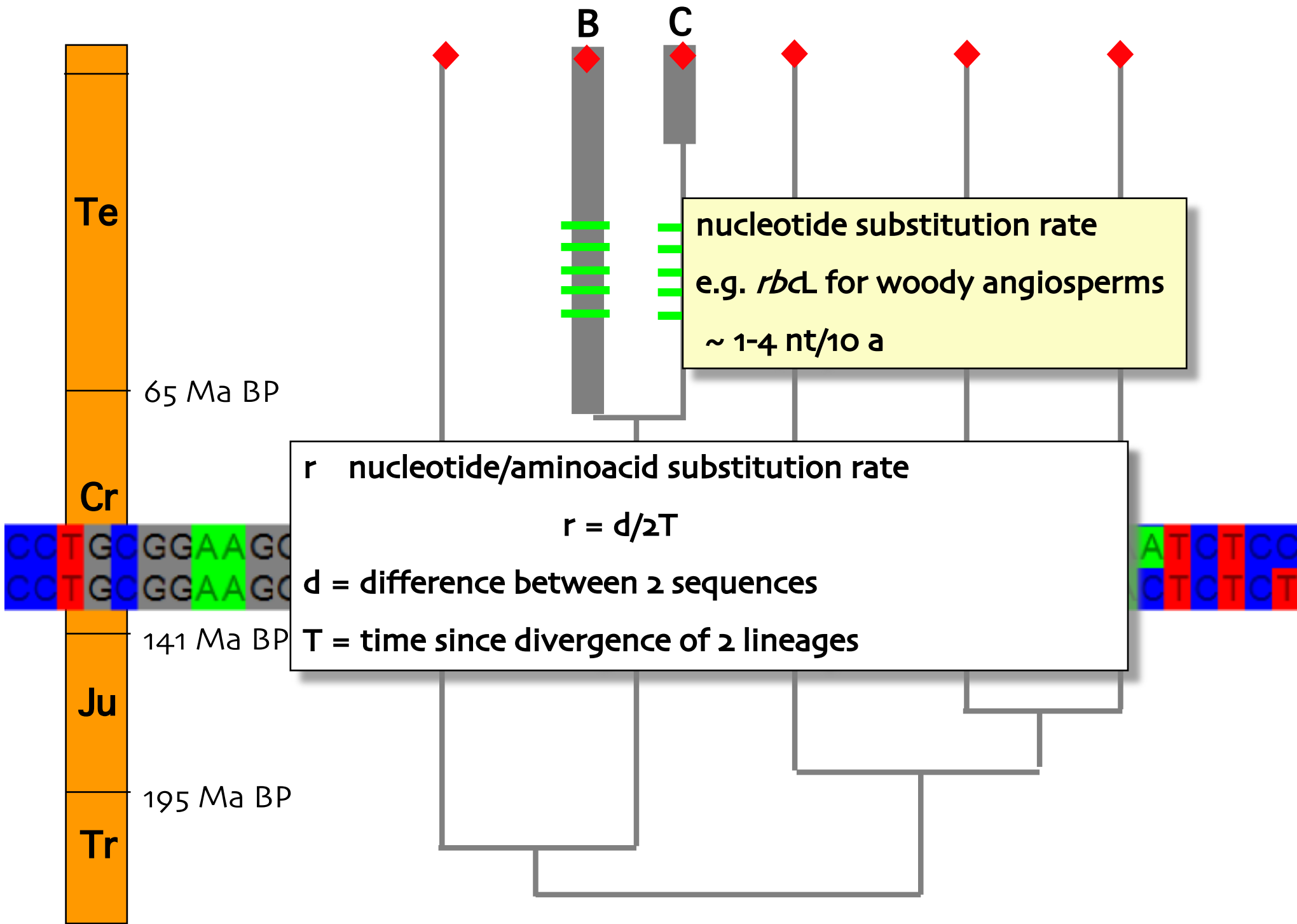
195 Ma BP

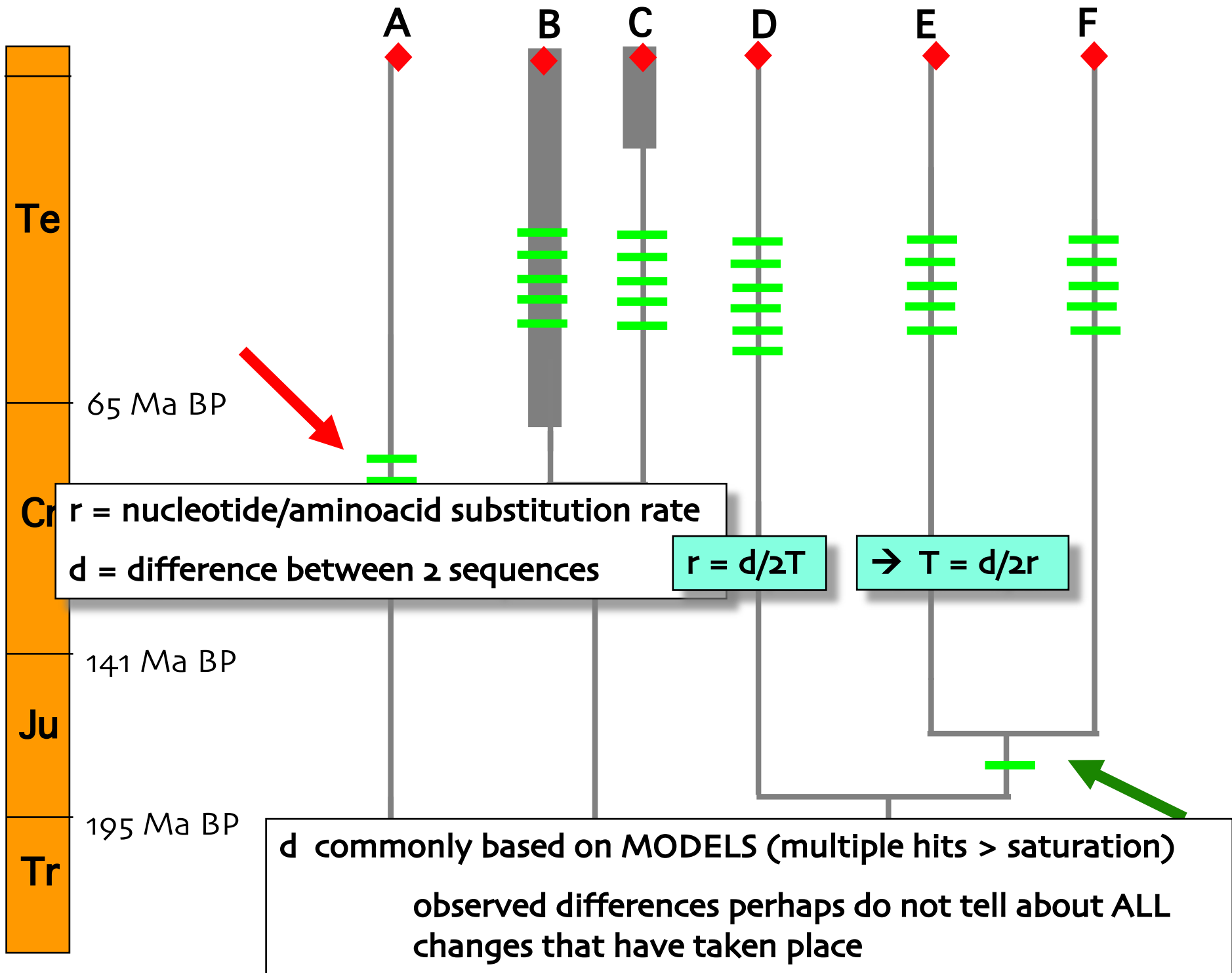
Tr



FOSSILS IN CLADISTIC ANALYSIS

minimum age for clades
calibration of molecular "clock"



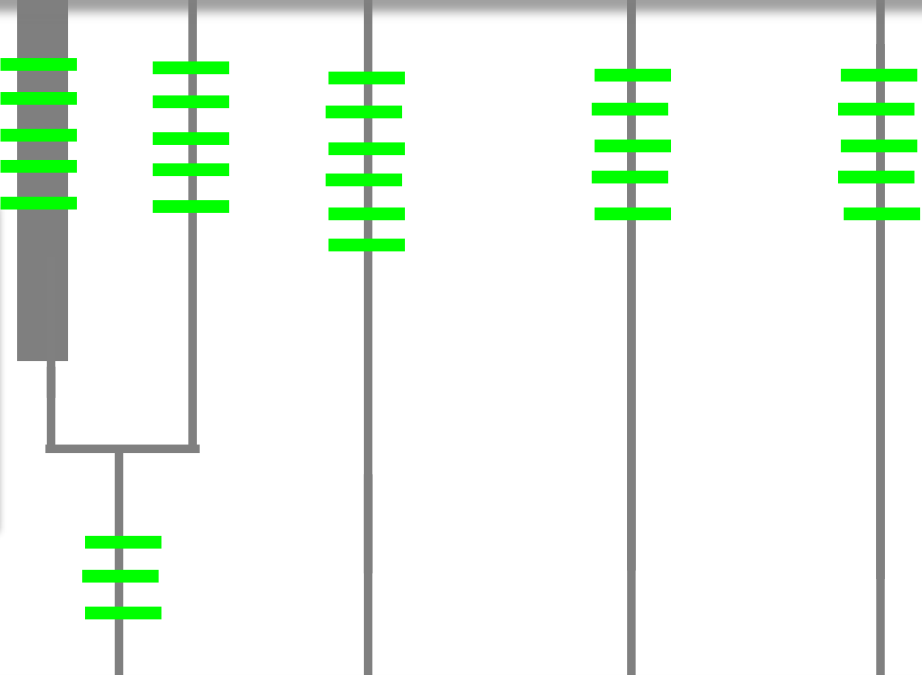


A B C D E F

Crepet, W.L. & al. 2004. Fossil evidence and phylogeny: the age of major angiosperm clades based on mesofossil and macrofossil evidence from Cretaceous deposits. *American Journal of Botany* 91: 1666-1682

Te

it is of utmost importance that we can be confident about MONOPHYLY of studied lineages



Ju

141 Ma BP

i.e. that the key fossil REALLY belongs to a certain lineage, it possess synapomorphies of this lineage

Tr

195 Ma BP

unfortunately this basic requirement is often IGNORED



FOSSILS IN CLADISTIC ANALYSIS

Graur, D. & Martin, W. 2004. Reading the entrails of chickens: molecular timescales of evolution and the illusion of precision. *Trends in Genetics* 20: 80-86

minimum age for clades

calibration of molecular "clock"

LOCAL

HETEROTACHY

variation of substitution rates in different lineages

constraints on nucleotide composition, e.g. genes
coding for RNA, secondary- & tertiary structures

UNKNOWN fossils (e.g. herbaceous conifers!)

HUGE variation in possible values for the ages of lineages

Gallego, J. & Arias, J.S. 2012. Are calibration methods stable? An empirical study using penalized likelihood. *XXXI Willi Hennig Meeting. Abstracts*

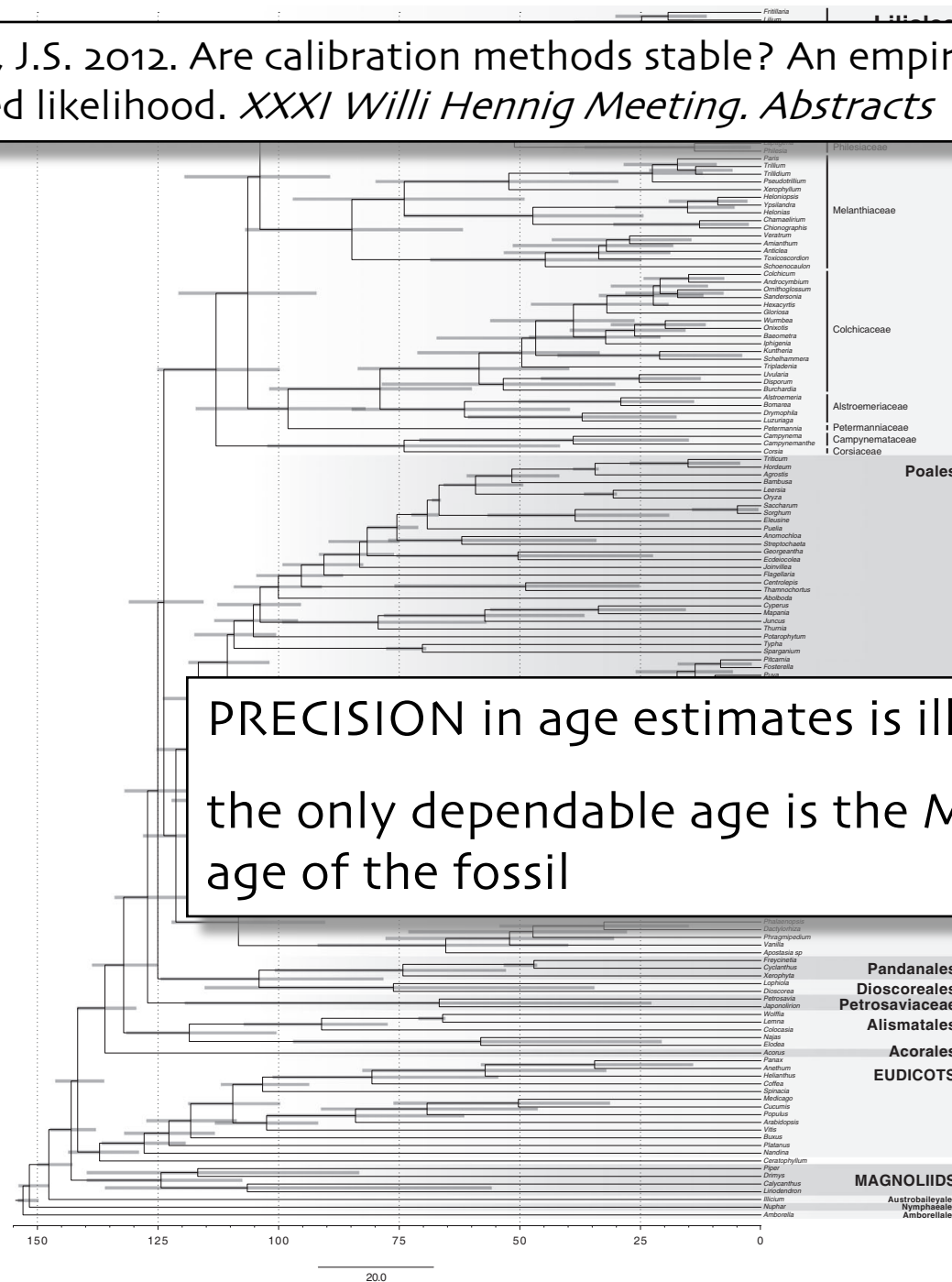
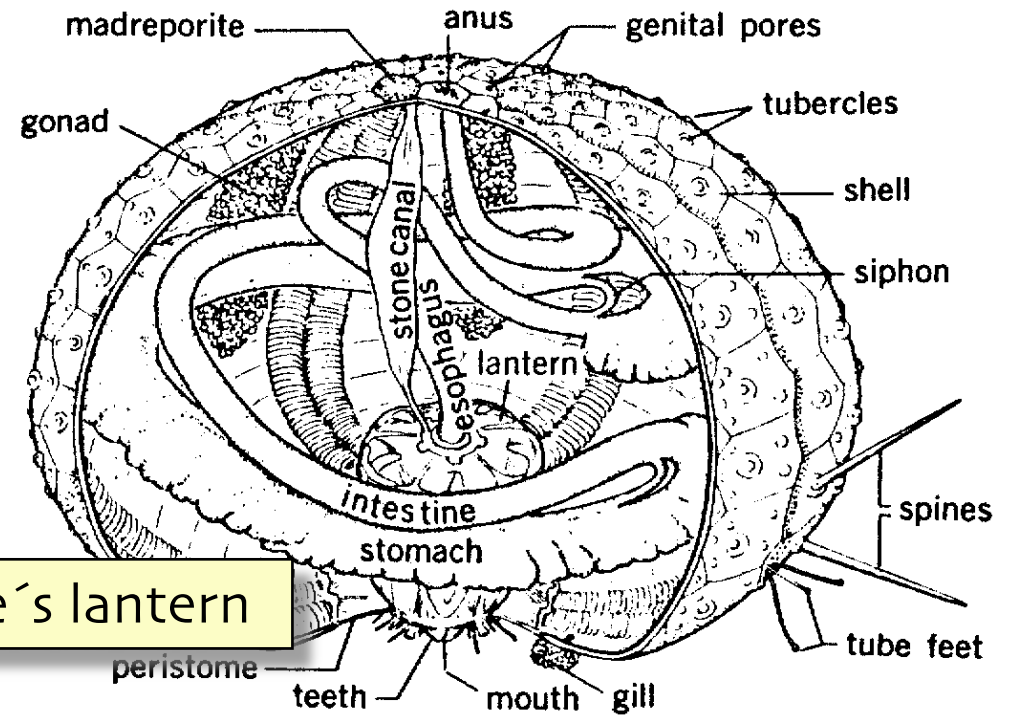
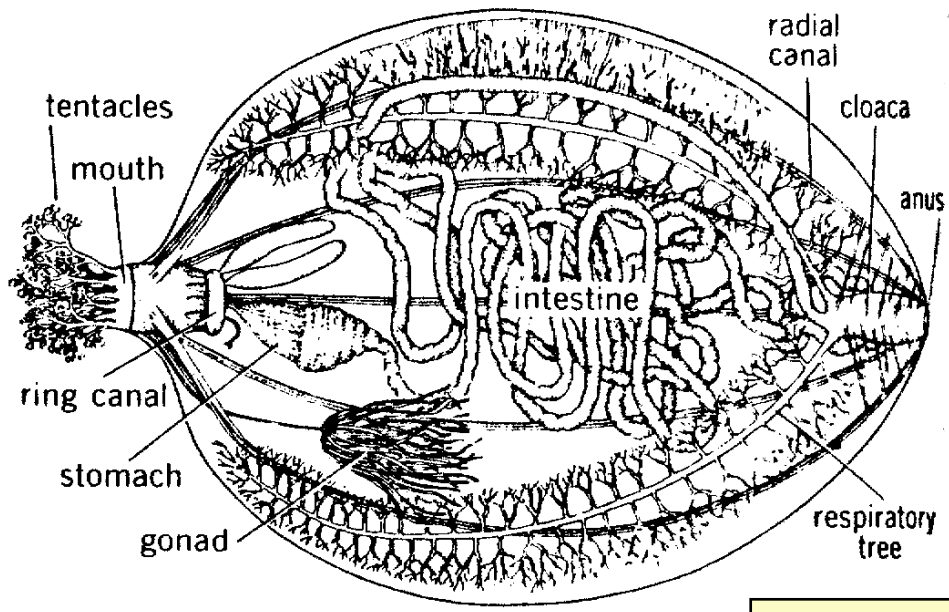


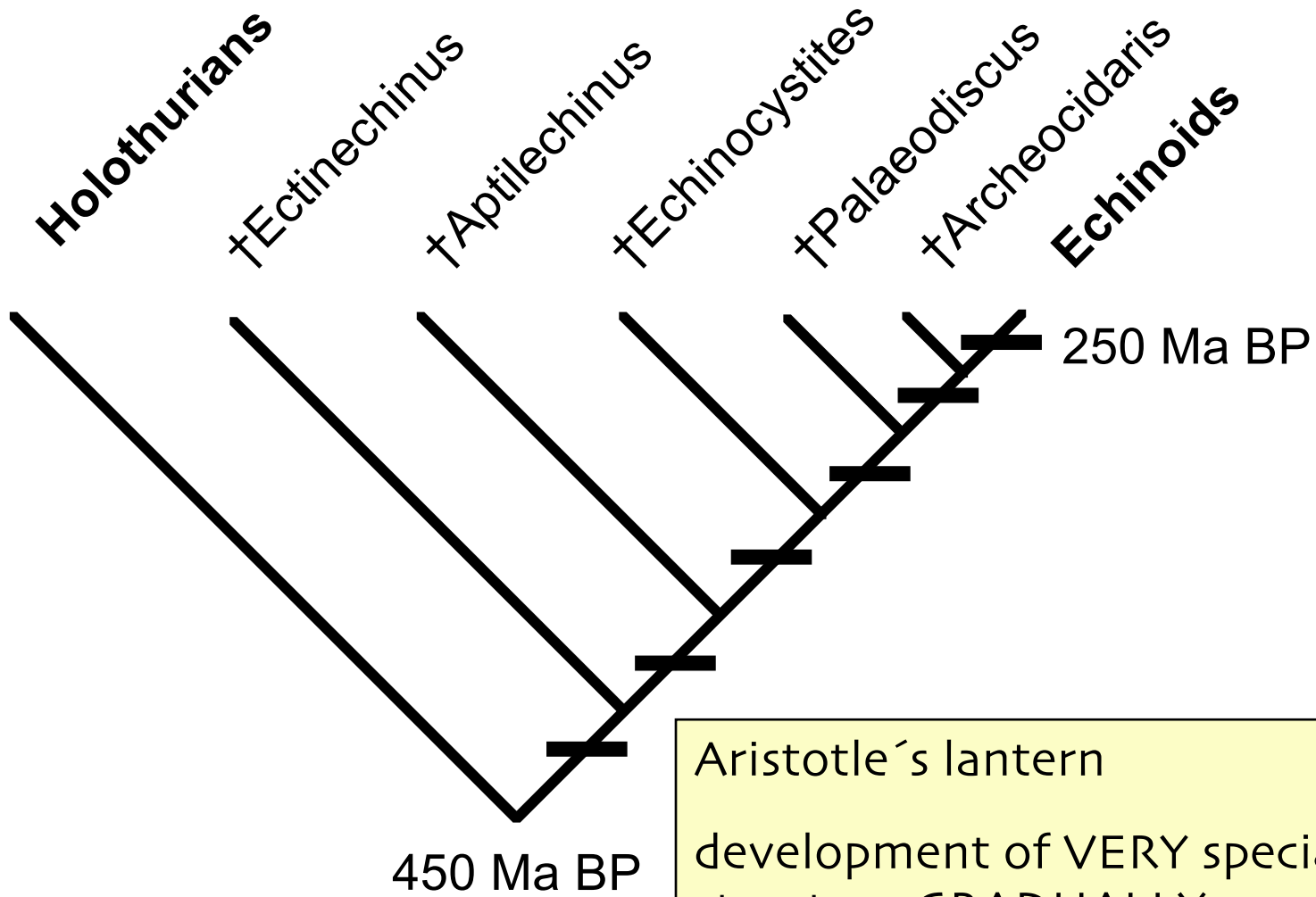
Fig. 3. Simplified timeline in millions of years for monocot evolution based on BEAST analysis, including single place-holders for each genus; grey bars represent 95% higher probability densities around each mean.

FOSSILS IN CLADISTIC ANALYSIS

minimum age for clades
calibration of molecular "clock"
assumptions about homology



Aristotle's lantern

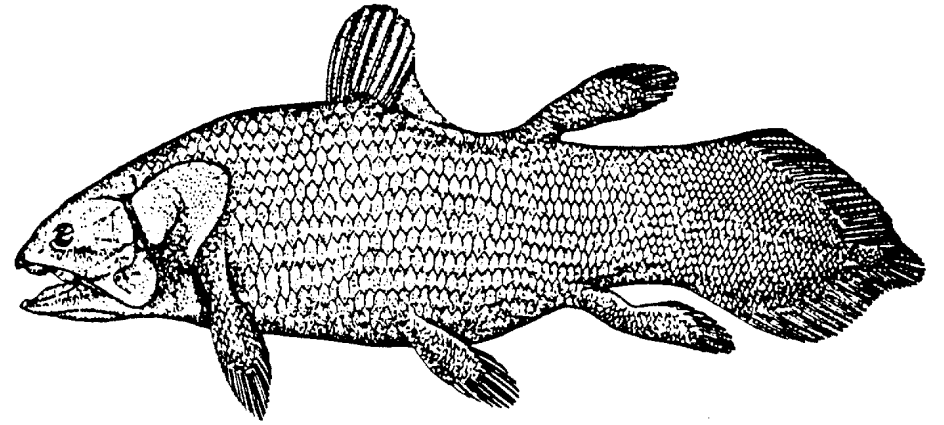
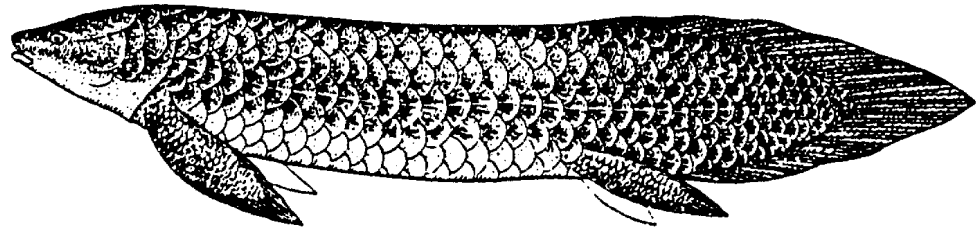


FOSSILS IN CLADISTIC ANALYSIS

minimum age for clades
calibration of molecular "clock"

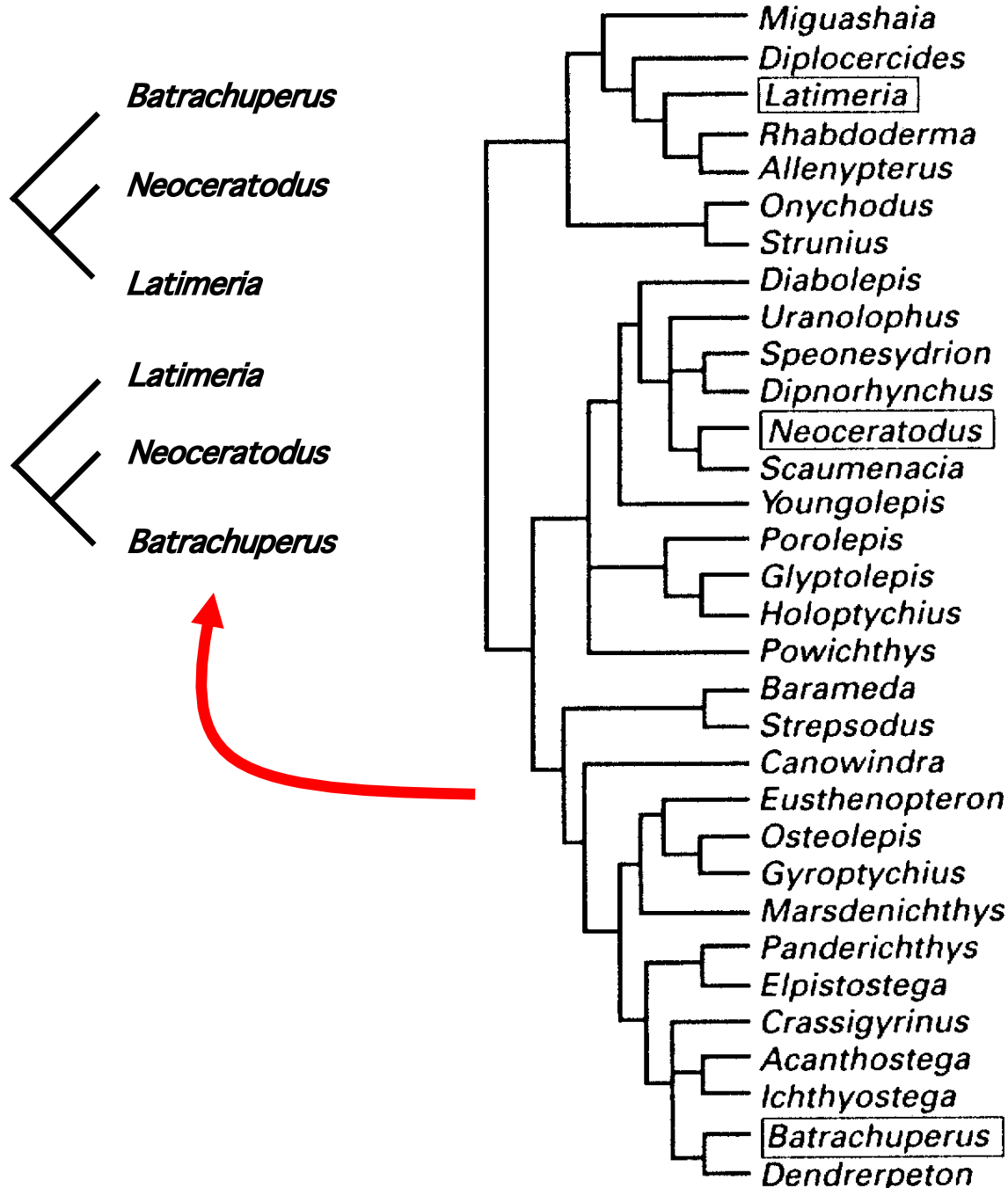
assumption about homology

fossils can alter results of WHOLE analysis



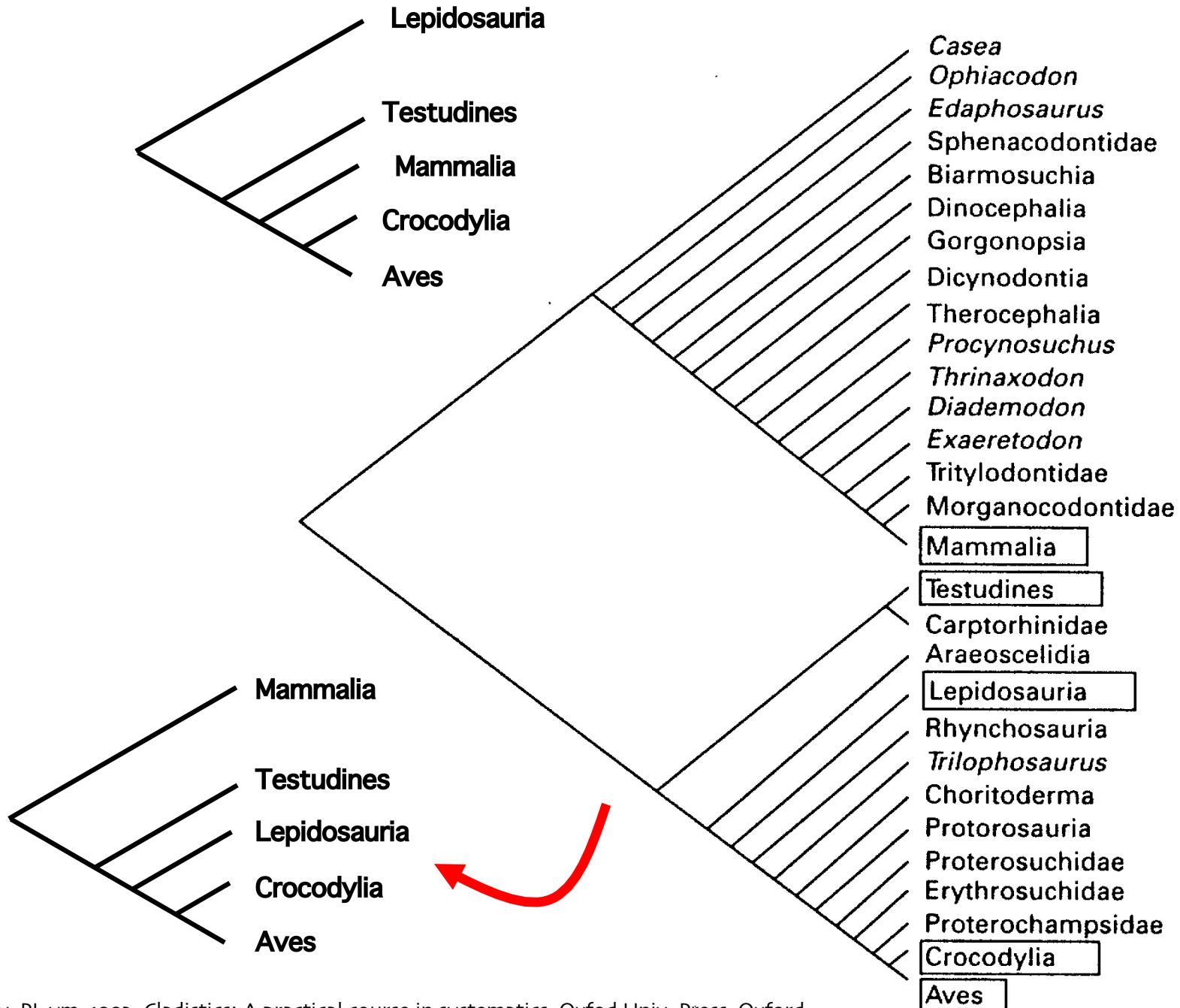
extant taxa

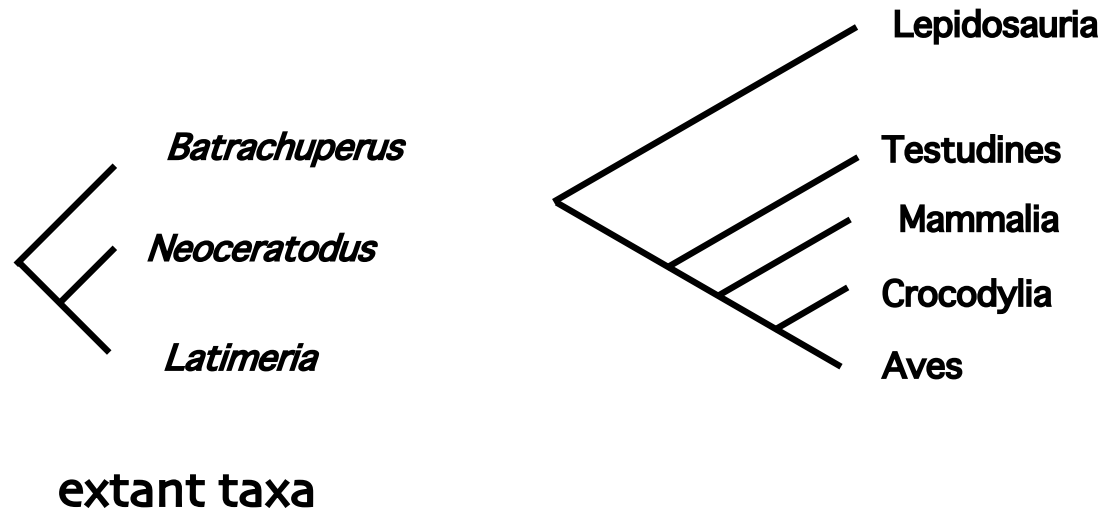
fossils included



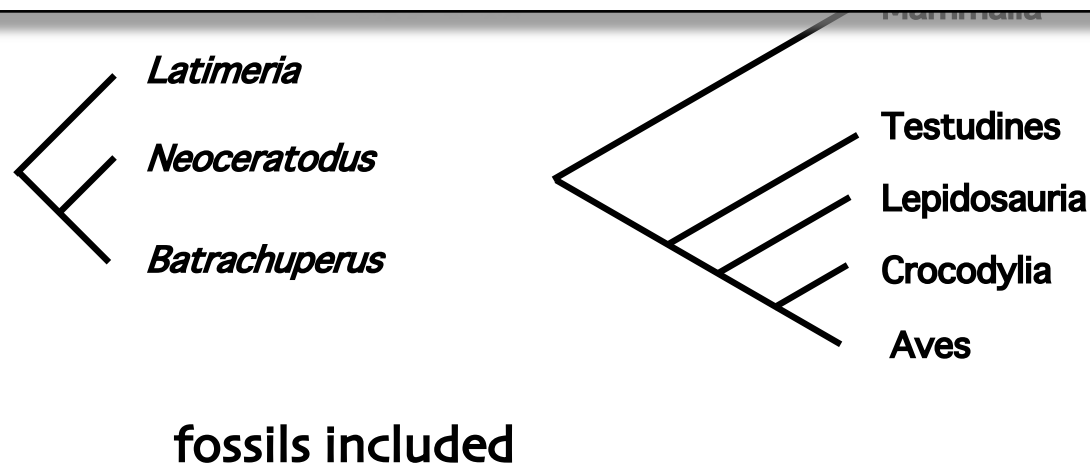
extant taxa

fossils included





Rothwell, G.W. & Nixon, K.C. 2006. How does inclusion of fossil data change our conclusions about the phylogenetic history of euphylllophytes?
Int. J. Plant Sci. 167: 737-749.



FOSSILS IN CLADISTIC ANALYSIS

minimum age for clades
calibration of molecular "clock"

assumptions about homology

fossils can alter results of WHOLE analysis
often LOTS of missing information

small amount of information

FOSSILS IN CLADISTIC ANALYSIS



A	0100011100	1010110101	10001110100	01110111111100011	1111110111101
B	0111110000	1001001001	10110010001	10111111111011001	1001010001000
C	010110100	10001111010	0110011111	1100011101100000	1111111001010
†D	?		0011000101	?	
E	010010100	01101111000	0100001000	100010100100000	01101101111101
†F	?	10101100010	1011100110	?	1001110101100
	pollen	stem	lvs	fruit	seed

FOSSILS IN CLADISTIC ANALYSIS

A	0100011100	1010110101	0001110100	0111011111100011	1111110111101
B	0111110000	1001001001	0110010001	10111111111011001	1001010001000
C	010110100	10001111010	0110011111	1100011101100000	1111111001010
†D	0 11	10 111 0	0011000101	0 11 10	111 0
E	010010100	01101111000	0100001000	1000101001000001	01101101111101
†F	?	10101100010	1011100110	?	1001110101100
	pollen	stem	lvs	fruit	seed

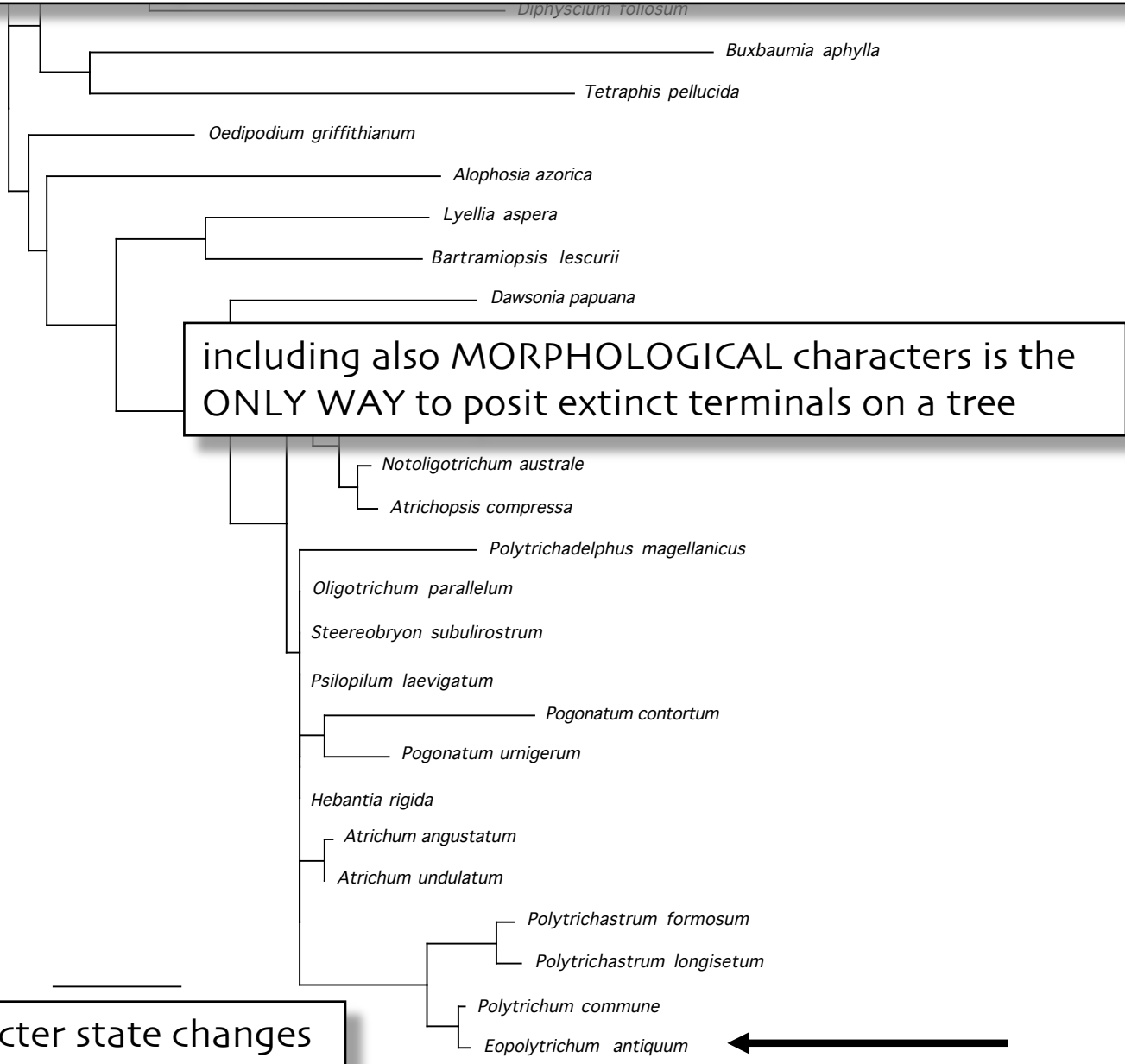
SIMULTANEOUS inclusion of ALL characters ensures reliable analysis & positing of terminals on tree

particularly IMPORTANT in studies of FOSSILS

WHY?

Sphagnum palustre

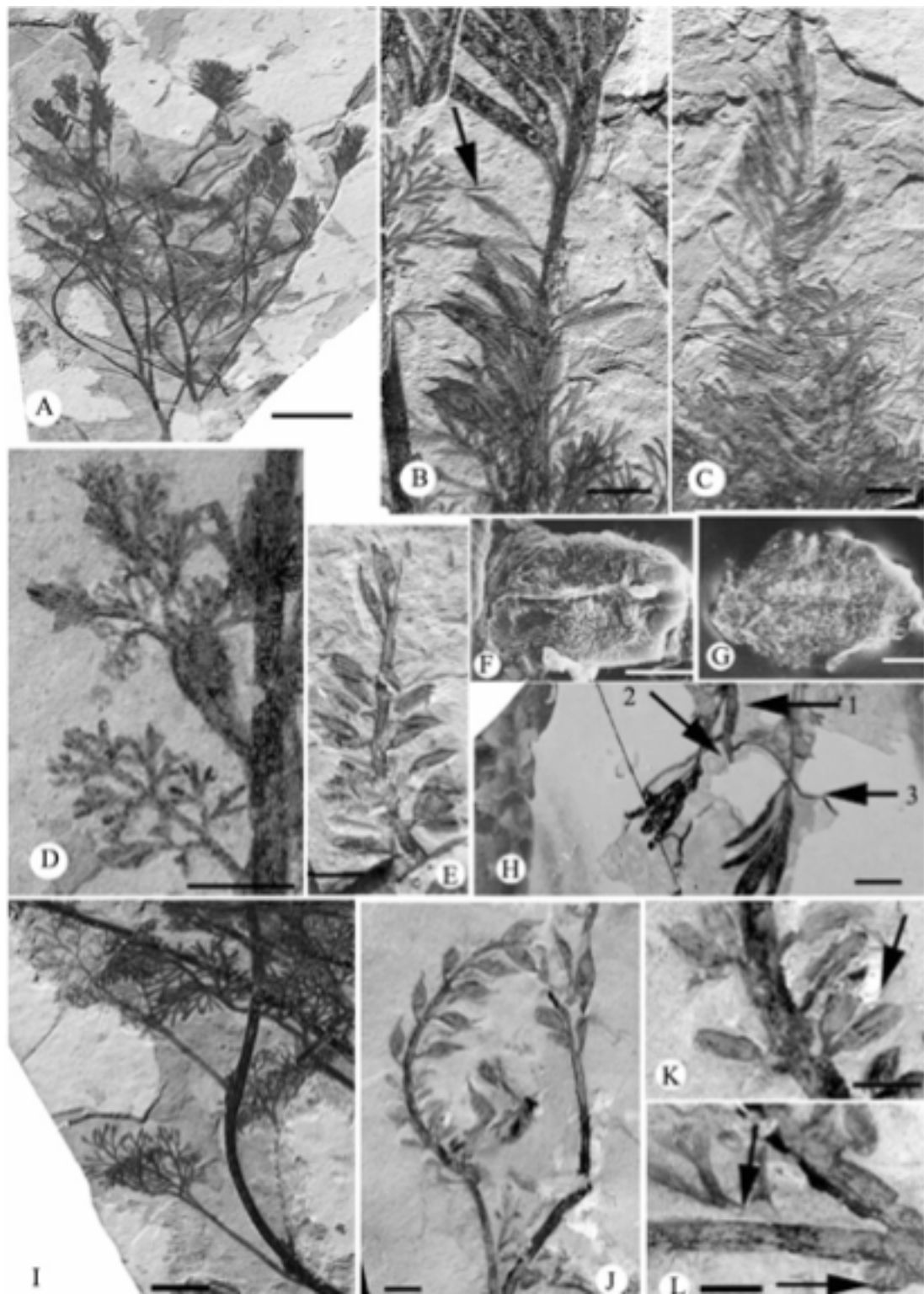
Bippus, A.C. & al. 2018. Wanted dead or alive (probably dead): stem group Polytrichaceae. *Am. J. Bot.* 105: 1-21.

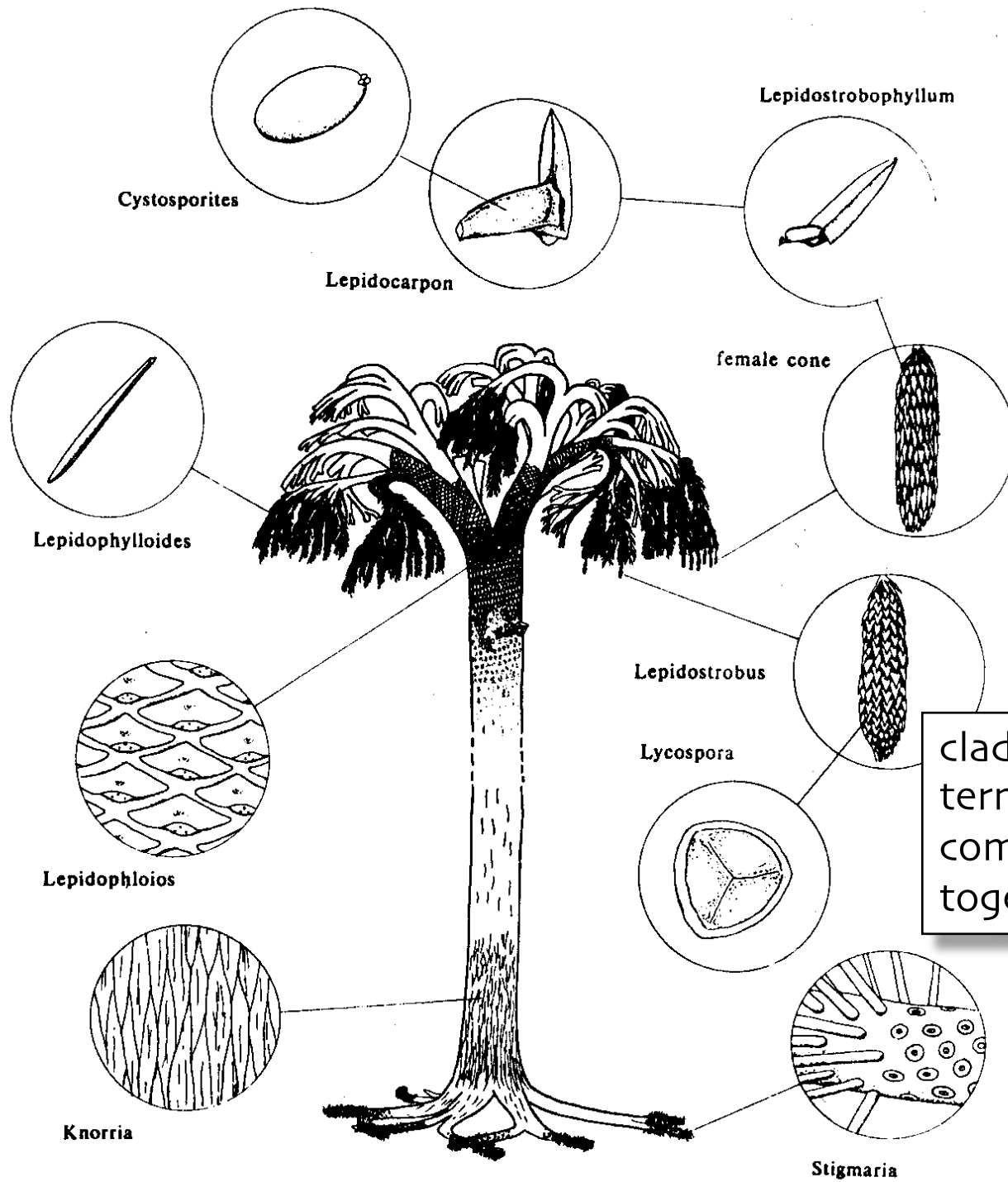


including also MORPHOLOGICAL characters is the ONLY WAY to posit extinct terminals on a tree

20 character state changes

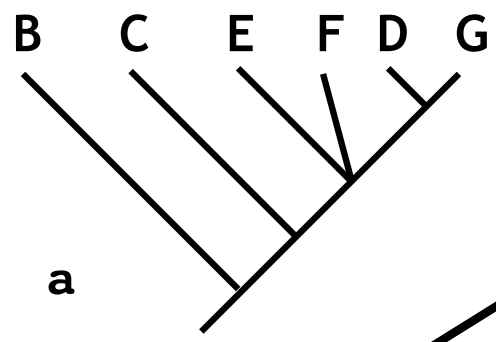




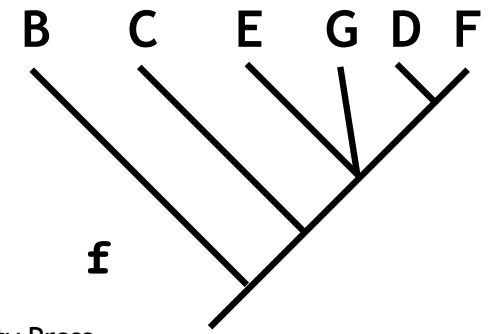
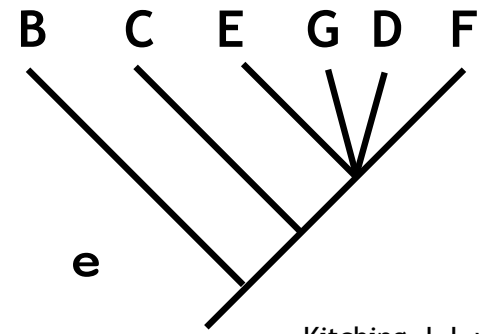
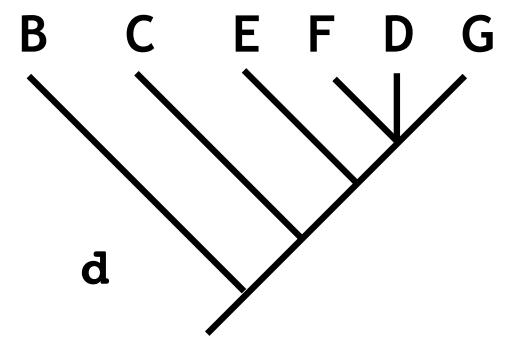
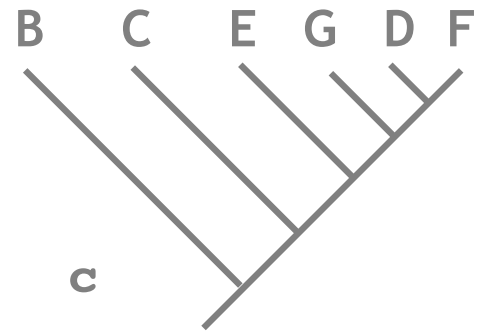
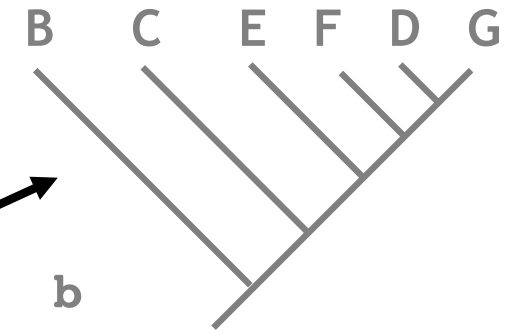


cladistic analysis with extant terminals MIGHT assist in combining fragmentary fossils together

		characters			
		0	1	2	3
terminals	A	0	0	0	0
	B	1	0	0	1
	C	1	1	0	0
	D	1	1	1	1
	E	1	1	1	0
	F	1	1	1	?
	G	1	1	1	?



IMPOSSIBLE trees
in terms of matrix



A
B
C
D
E
F
G

2 3

0 0

0 1

0 0

1 1

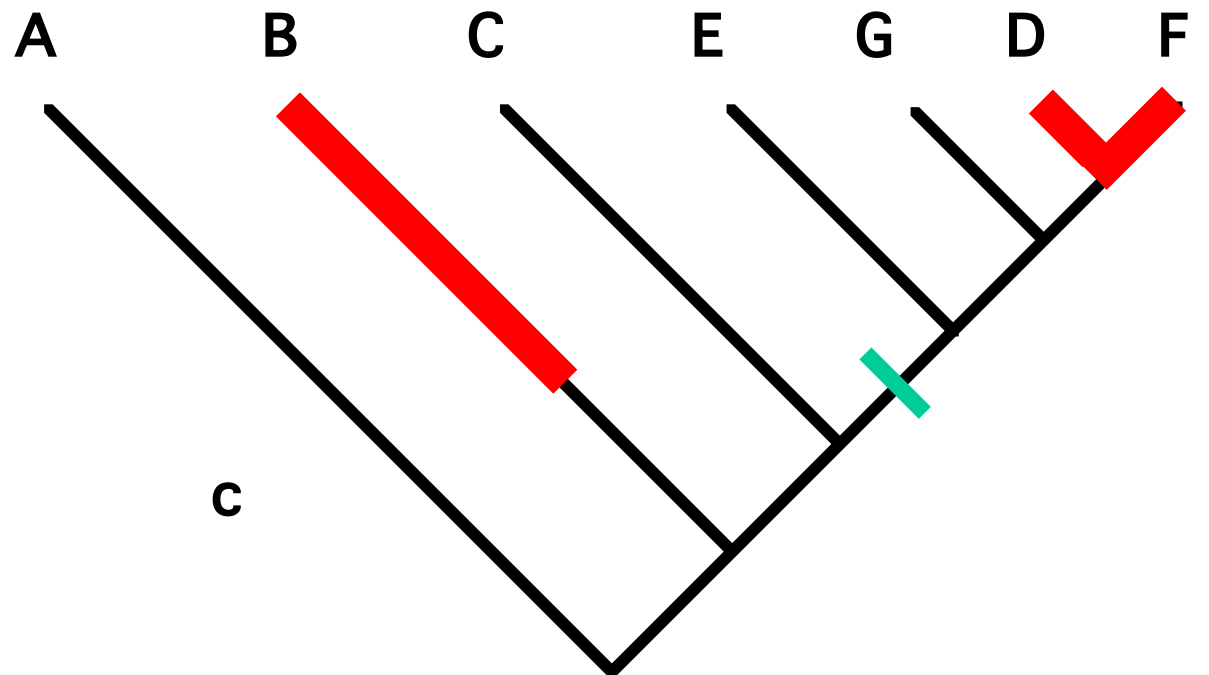
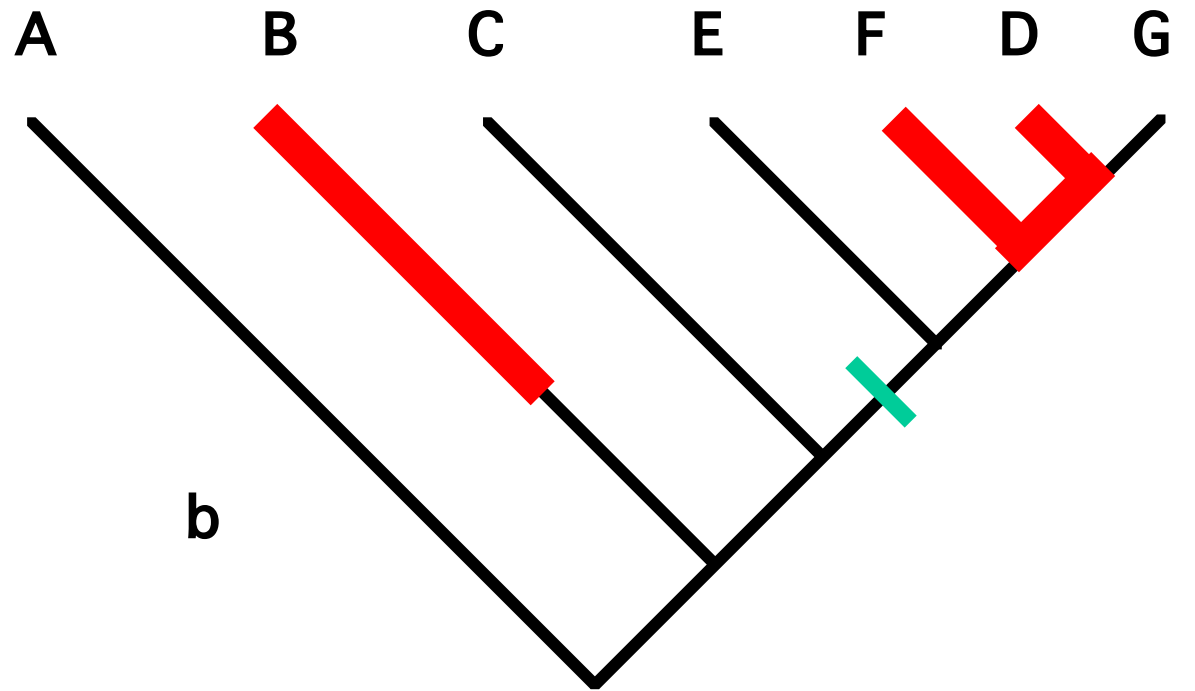
1 0

1 ? 0 1 1 0

1 ? 0 1 0 1

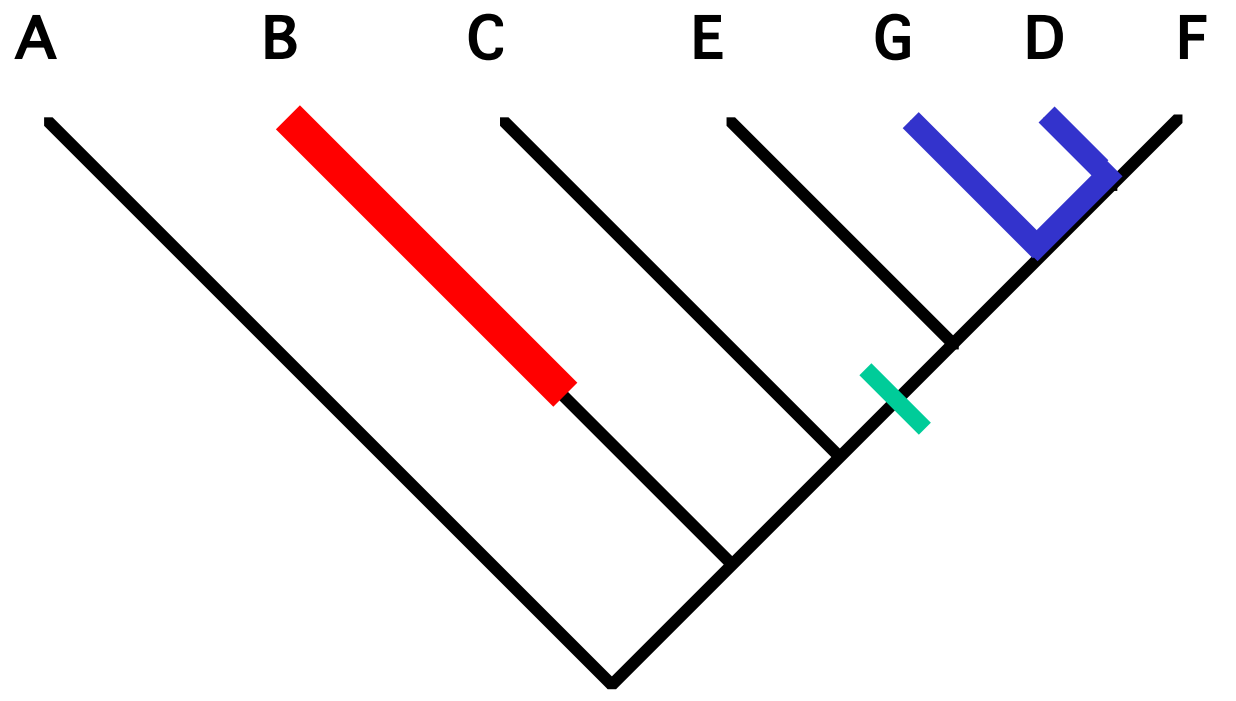
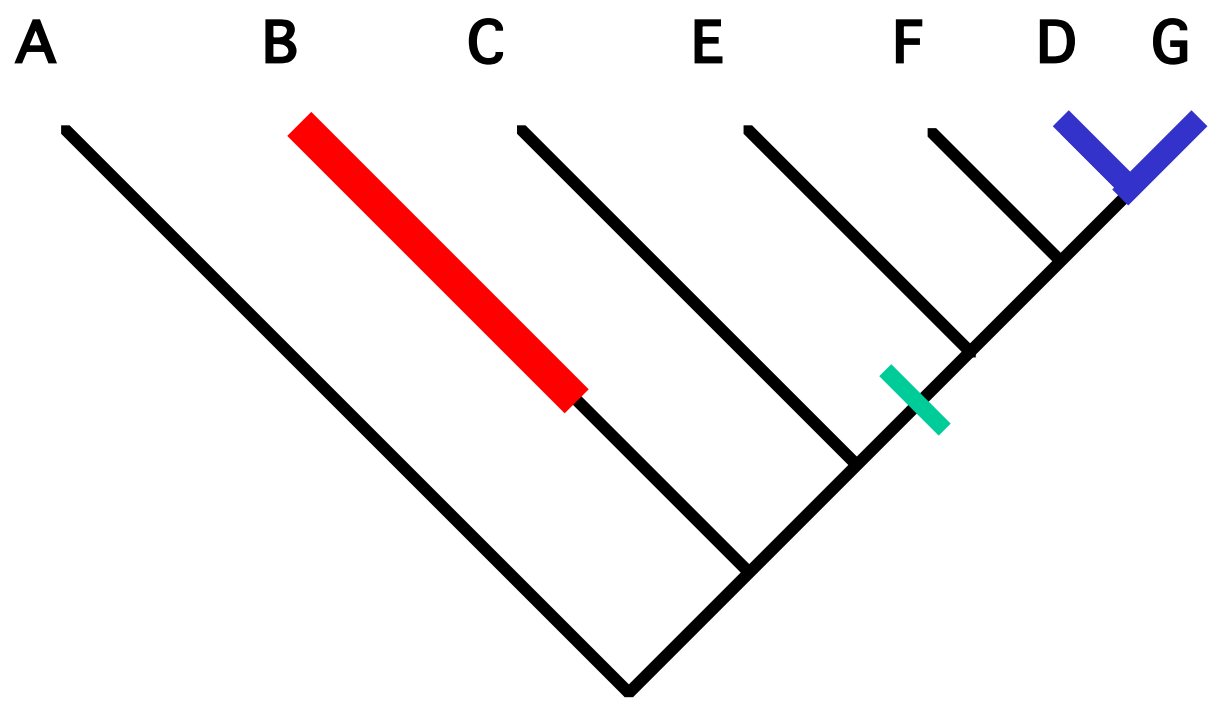
missing values are
treated as

**MAXIMALLY
POLYMORPHIC**



A
B
C
D
E
F
G

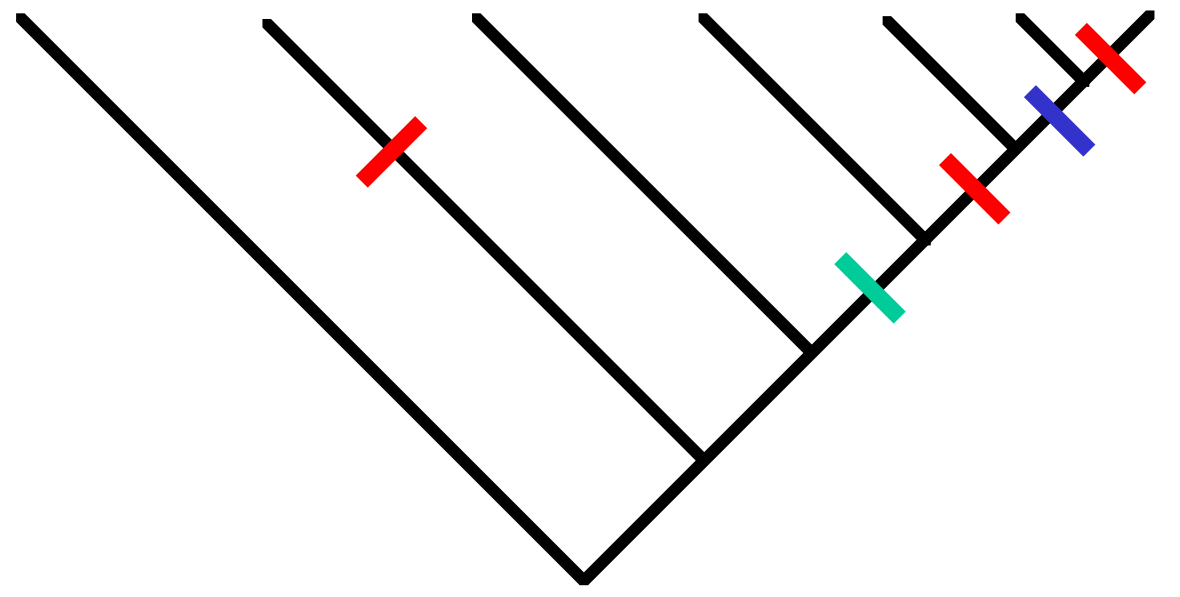
0	0				
0	1				
0	0				
1	1				
1	0				
1	?	0	1	1	0
1	?	0	1	0	1



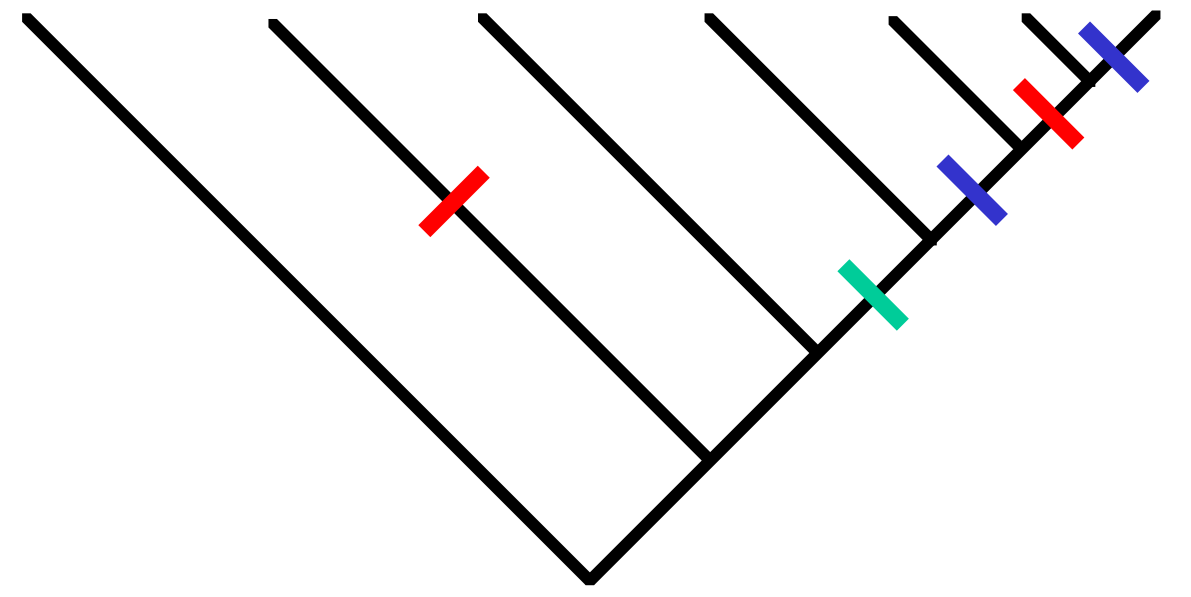
A
B
C
D
E
F
G

0	0				
0	1				
0	0				
1	1				
1	0				
1	?	0	1	1	0
1	?	0	1	0	1

A B C E F D G

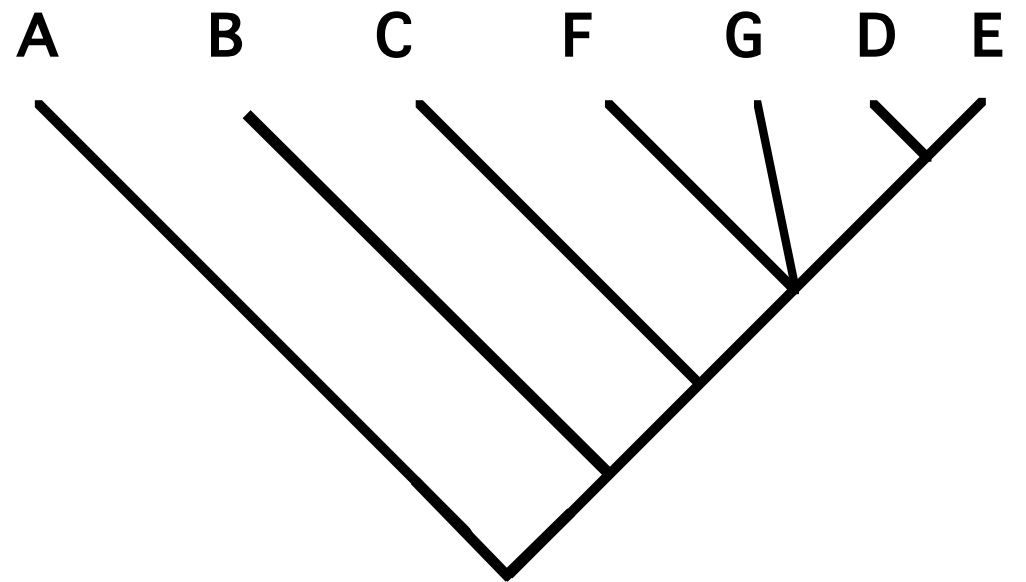
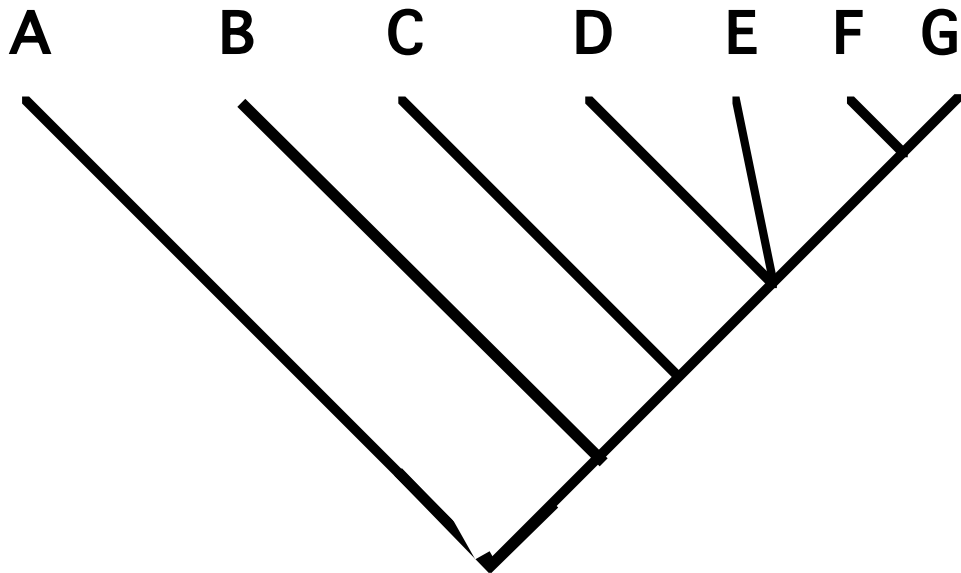
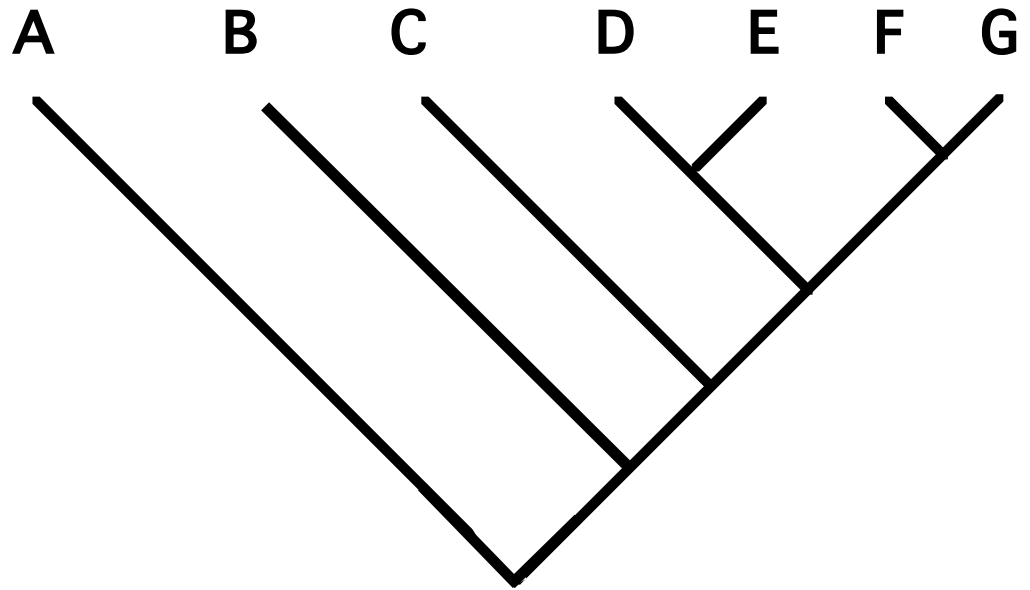


A B C E G D F

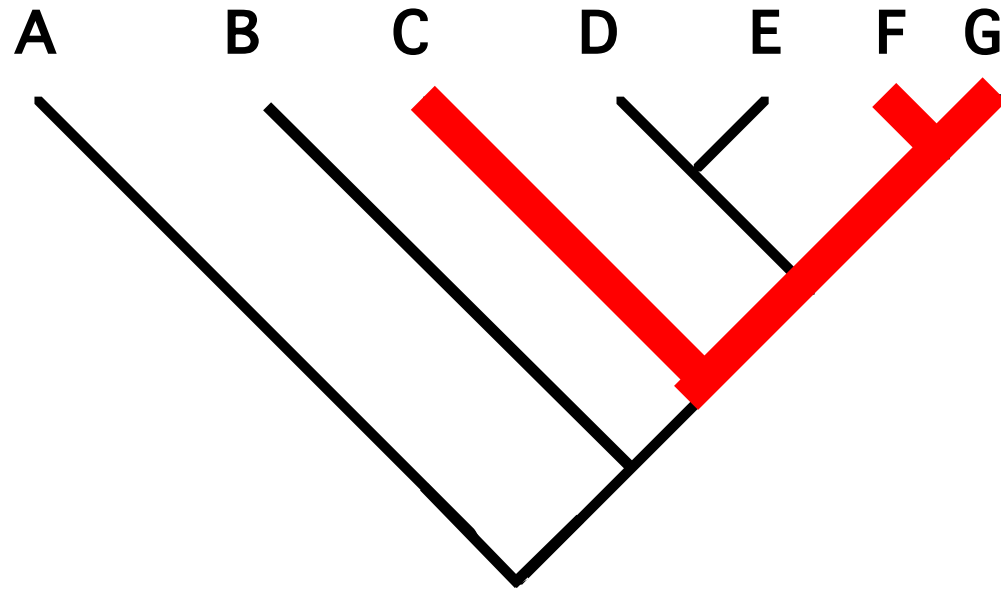


1 2 3 4 5

A	0	0	0	0	0
B	0	0	0	0	1
C	1	0	0	1	1
D	0	1	1	1	1
E	0	1	1	1	1
F	1	1	1	1	1
G	1	1	1	1	1

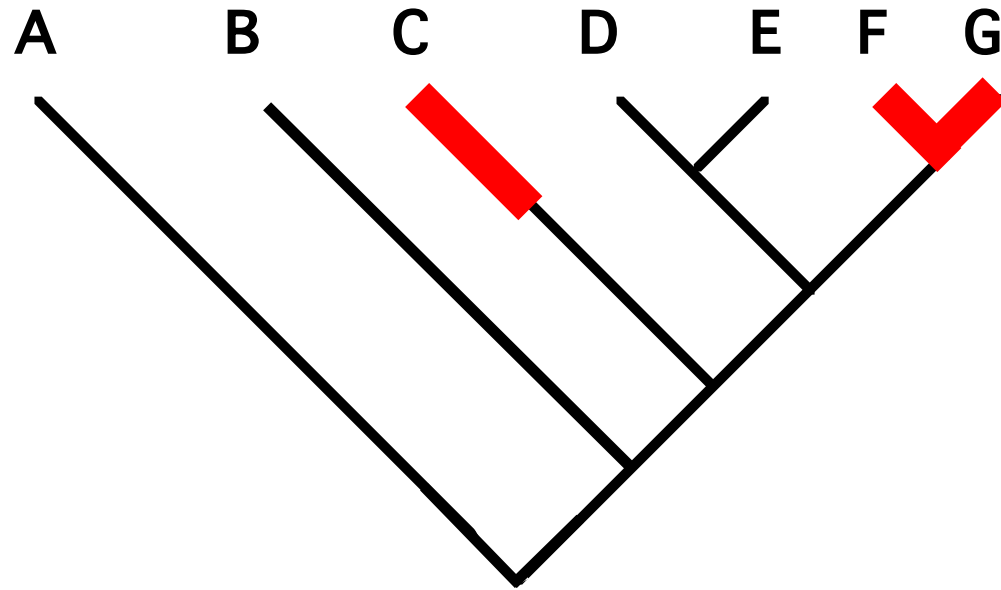


	1	2	3	4	5
A	0	0	0	0	0
B	0	0	0	0	1
C	1	0	0	1	1
D	0	1	1	1	1
E	0	1	1	1	1
F	1	1	1	1	1
G	1	1	1	1	1



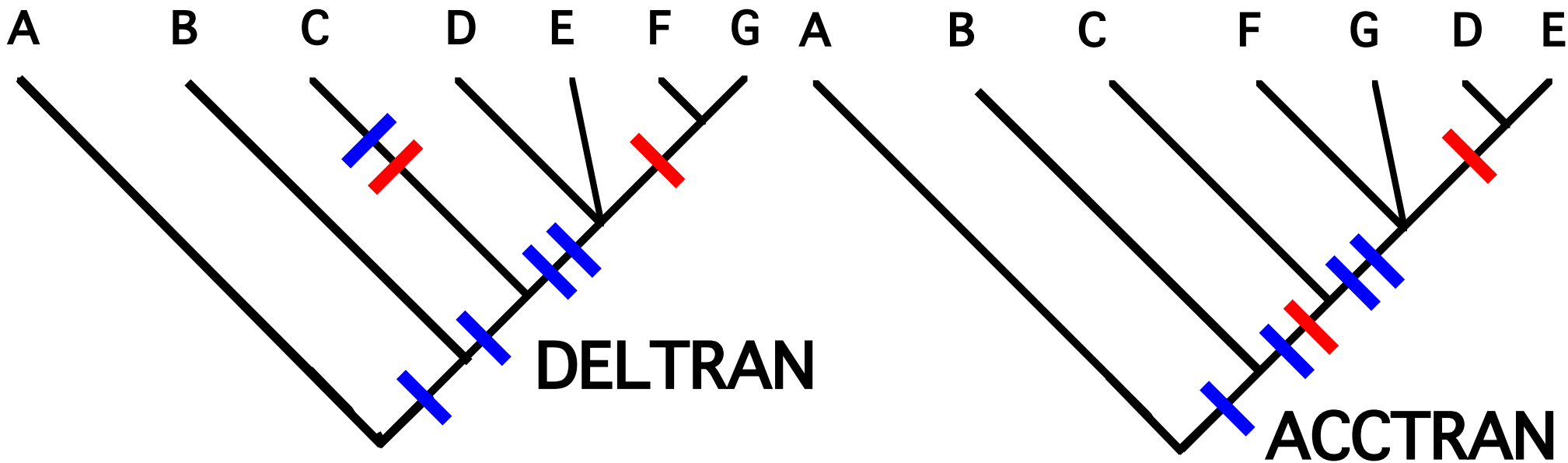
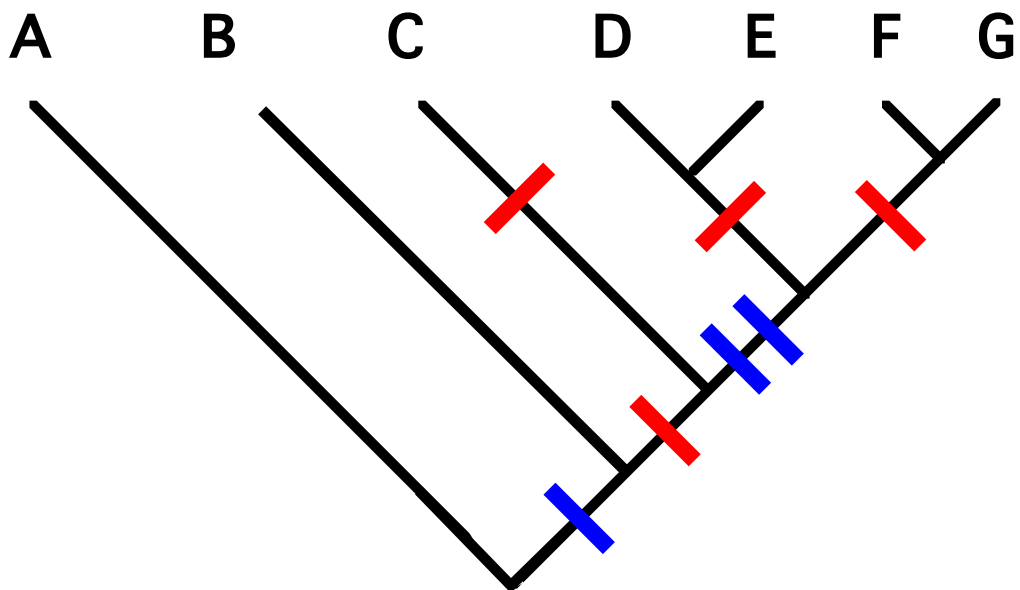
ACCTRAN optimization

	1	2	3	4	5
A	0	0	0	0	0
B	0	0	0	0	1
C	1	0	0	1	1
D	0	1	1	1	1
E	0	1	1	1	1
F	1	1	1	1	1
G	1	1	1	1	1



DELTRAN optimization

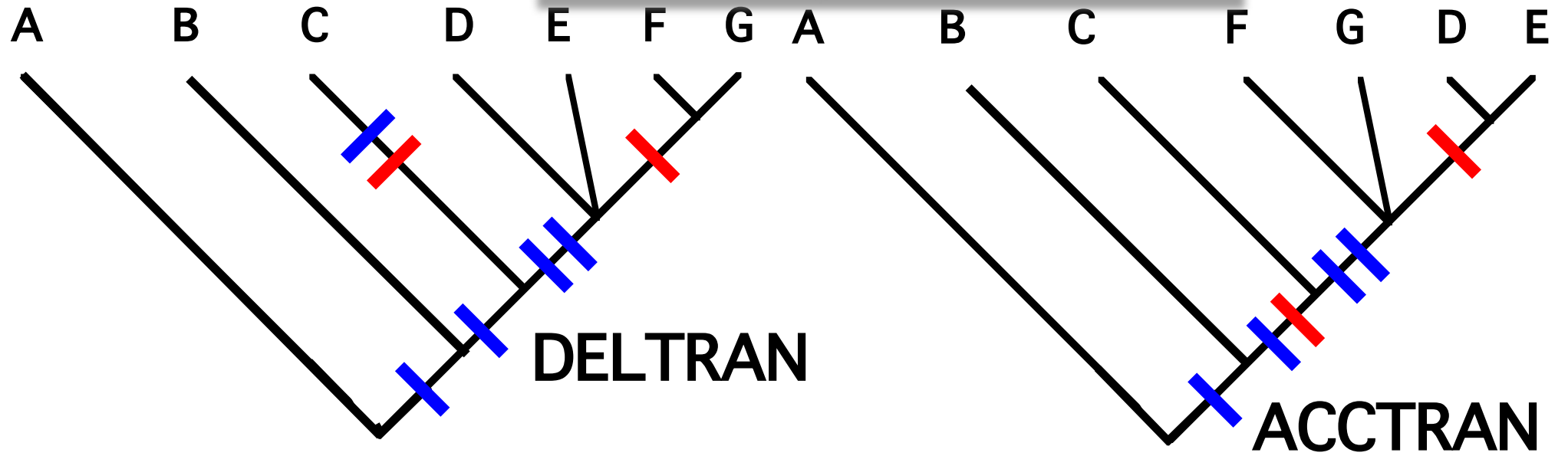
	1	2	3	4	5
A	0	0	0	0	0
B	0	0	0	0	1
C	1	0	0	1	1
D	0	1	1	1	1
E	0	1	1	1	1
F	1	1	1	1	1
G	1	1	1	1	1



	1	2	3	4	5
A	0	0	0	0	0
B	0	0	0	0	1
C	1	0	0	1	1
D	0	1	1	1	1
E	0	1	1	1	1
F	1	1	1	1	1
G	1	1	1	1	1

when some of equally parsimonious trees are MORE resolved than others SHOULD BE checked if all groups have synapomorphies SIMULTANEOUSLY

are some branches present only with some optimizations?



FOSSILS IN CLADISTIC ANALYSIS

minimum age for clades
calibration of molecular "clock"

homology

fossils can alter results of WHOLE analysis
often LOTS of missing information

small amount
of
information

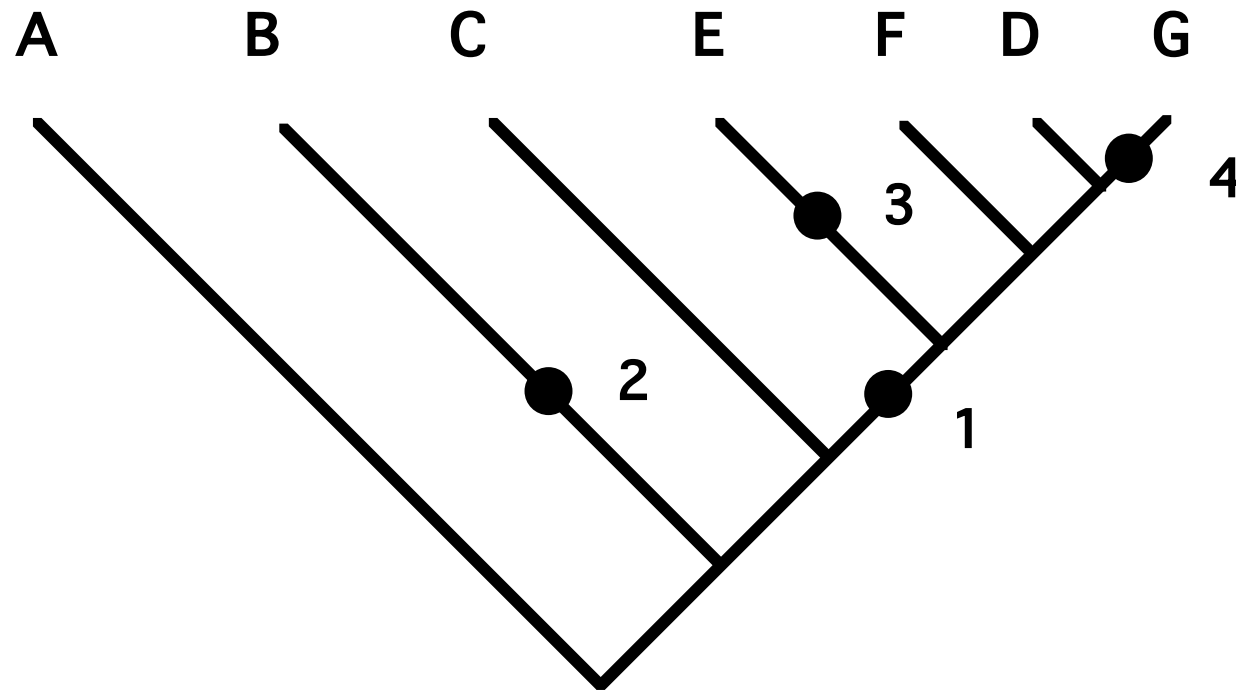
missing entries introduced also to EXTANT
terminals!

loss of resolution

possibly characters known only
from some extinct terminals

FOSSILS IN CLADISTIC ANALYSIS

fossils might alter **ALSO** position of other terminals included in analysis



● 1-4
alternative positions of fossil X

FOSSILS IN CLADISTIC ANALYSIS

minimum age for clades
calibration of molecular "clock"

homology

fossils can alter results of WHOLE analysis
often LOTS of missing information

small amount
of
information

Lundberg-rooting

Lundberg, J.G. 1972. Wagner networks and ancestors. *Systematic Zoology* 21: 398-413.

addition of fossils to analysis **AFTERWARDS**

local & **GLOBAL** parsimony

FOSSILS IN CLADISTIC ANALYSIS

minimum age for clades
calibration of molecular "clock"

assumptions about homology

fossils can alter results of WHOLE analysis
often LOTS of missing information

Lundberg-rooting

SAFE removal of terminals

FOSSILS IN CLADISTIC ANALYSIS

position of terminals DEFINED COMPLETELY by characters
WITHOUT missing information

algorithms optimize characters with missing information
as parsimoniously as possible

Wilkinson, M. 1995. Coping with abundant missing entries in phylogenetic inference using parsimony. *Systematic Biology* 44: 501-514.

when 2 terminals are identical in characters that are
parsimony informative & information about both
is available, the one with less information can be
left out without affecting result ---->

FOSSILS IN CLADISTIC ANALYSIS

position of terminals **DEFINED COMPLETELY** by
characters WITHOUT missing information

algorithms optimize characters with missing information
as parsimoniously as possible

Wilkinson, M. 1995. Coping with abundant missing entries in phylogenetic inference using parsimony. *Systematic Biology* 44: 501-514.

when 2 terminals are identical in characters that are
parsimony informative & information about both
is available the one with less information can be left
out without affecting result ---->

FOSSILS IN CLADISTIC ANALYSIS

Kearney, M. 2002. Fragmentary taxa, missing data, and ambiguity: mistaken assumptions and conclusions. *Systematic Biology* 51: 369-381.

A	2	0	1	2	0	1	0
B	?	0	?	?	0	?	0
C	0	?	1	2	?	?	?

reference terminal

can be removed

CANNOT be removed

based on REAL AVAILABLE character information A & C DIFFER

Wilkinson, M. & Benton, M.J. 1996. Sphenodontid phylogeny and the problems of multiple trees. *Phil. Trans. Royal Soc. London B*, 351: 1-16

originally > 20 000 equally parsimonious trees

safe removal of 6 terminals ---> only 2 trees

FOSSILS IN CLADISTIC ANALYSIS

minimum age for clades
calibration of molecular "clock"

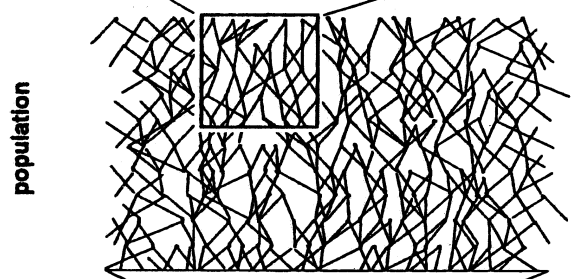
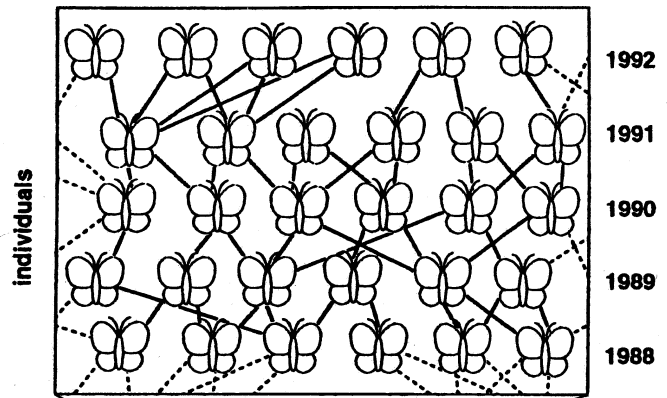
assumption about homology

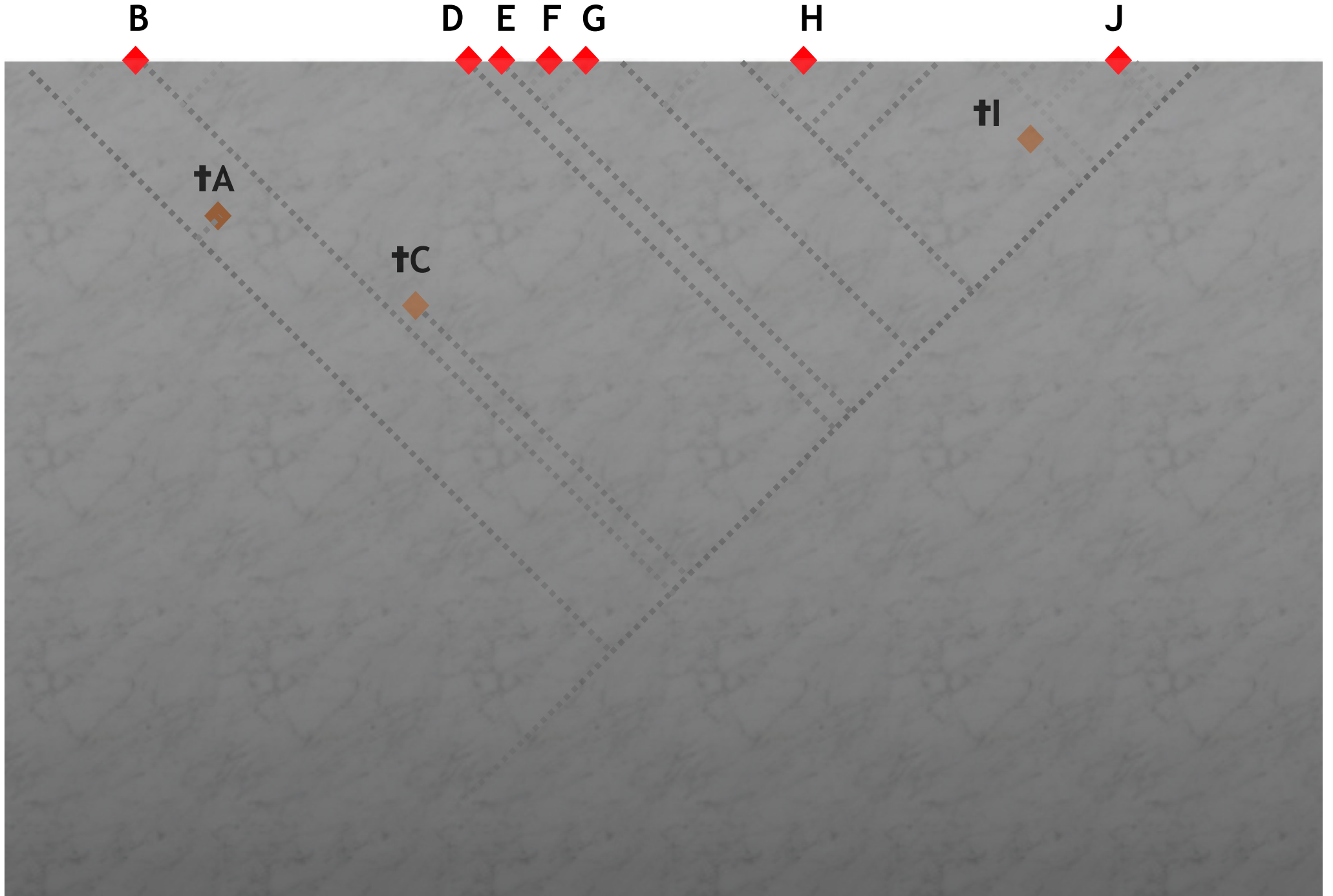
fossils can alter results of WHOLE analysis
often LOTS of missing information

Lundberg-rooting

SAFE removal of terminals

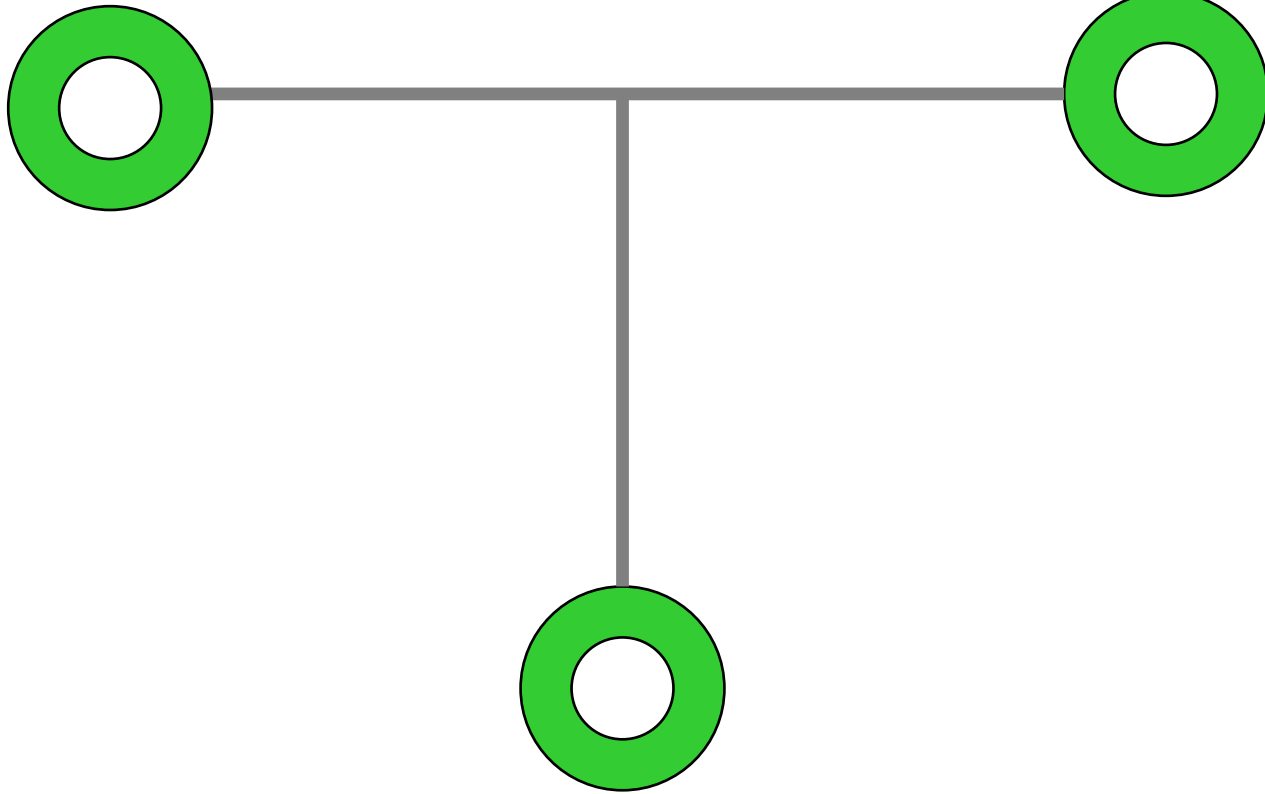
fossils as ancestors ?





A

B

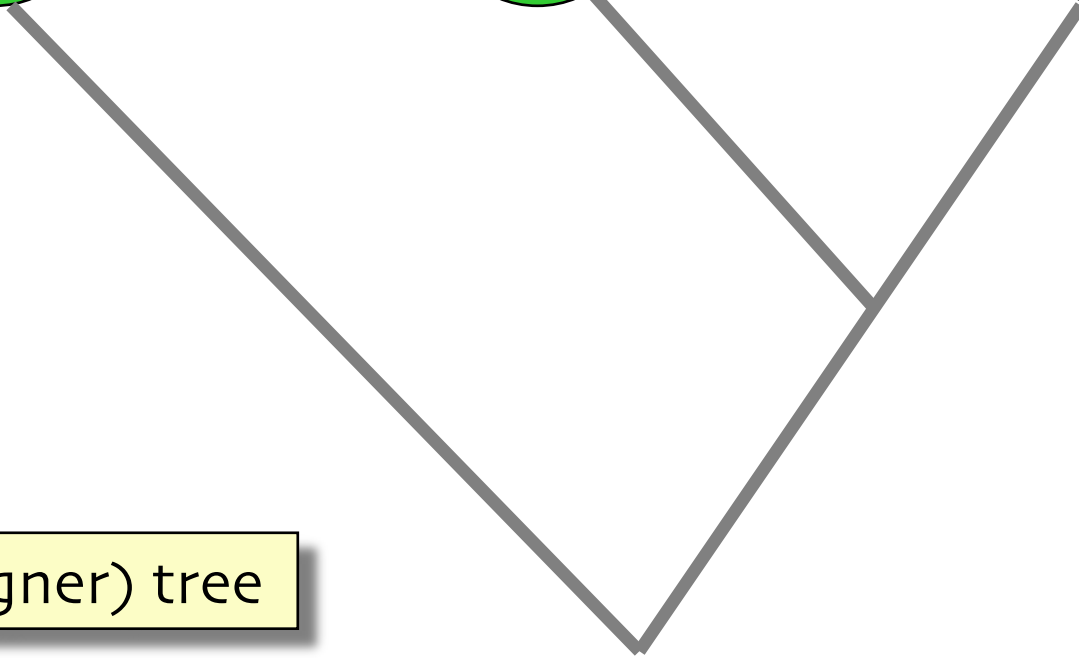
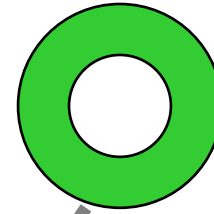
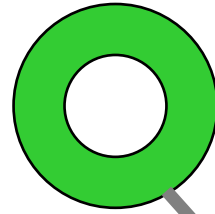
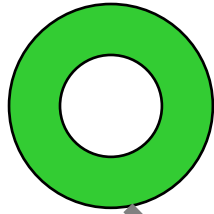


C

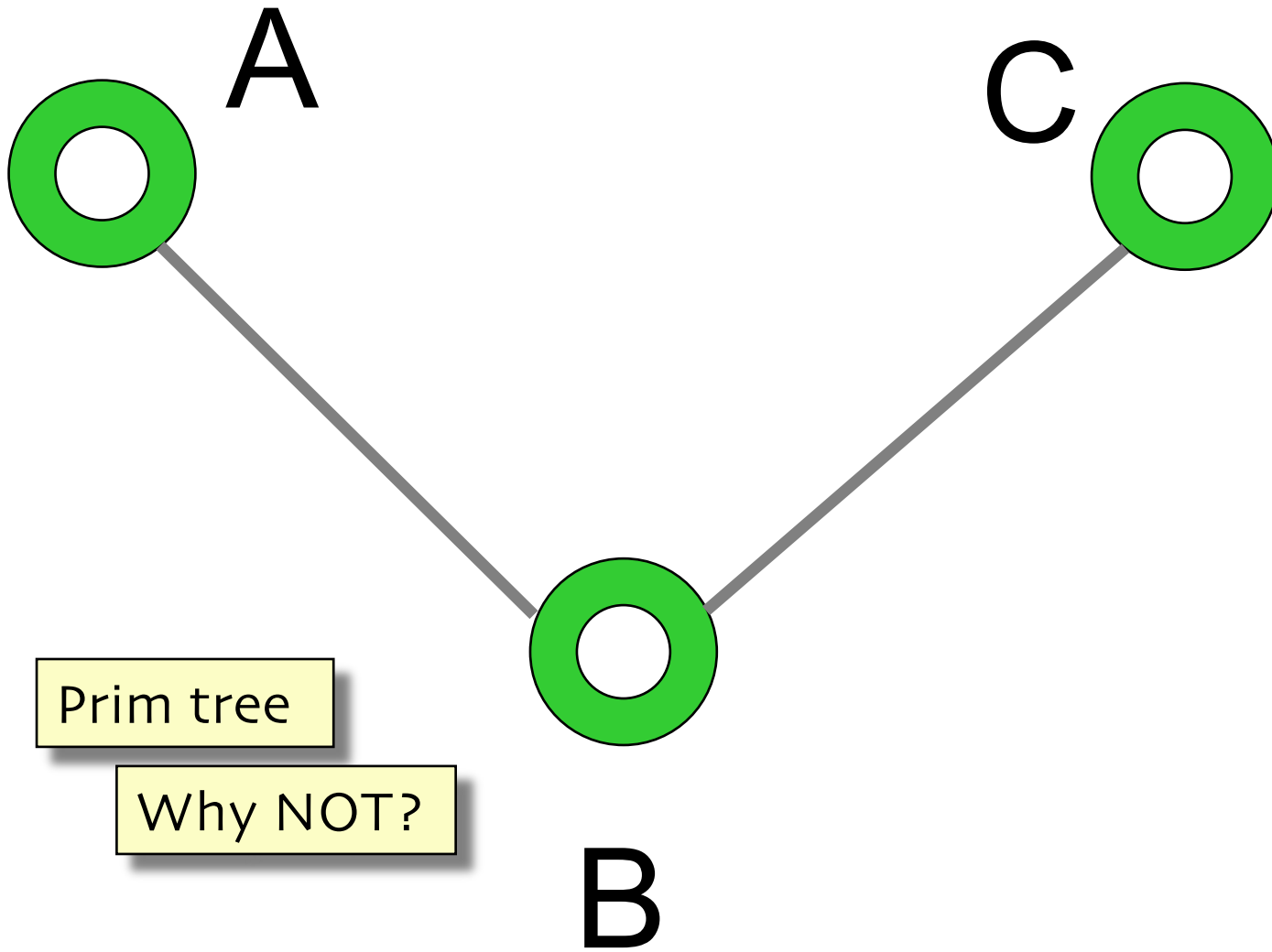
A

B

C



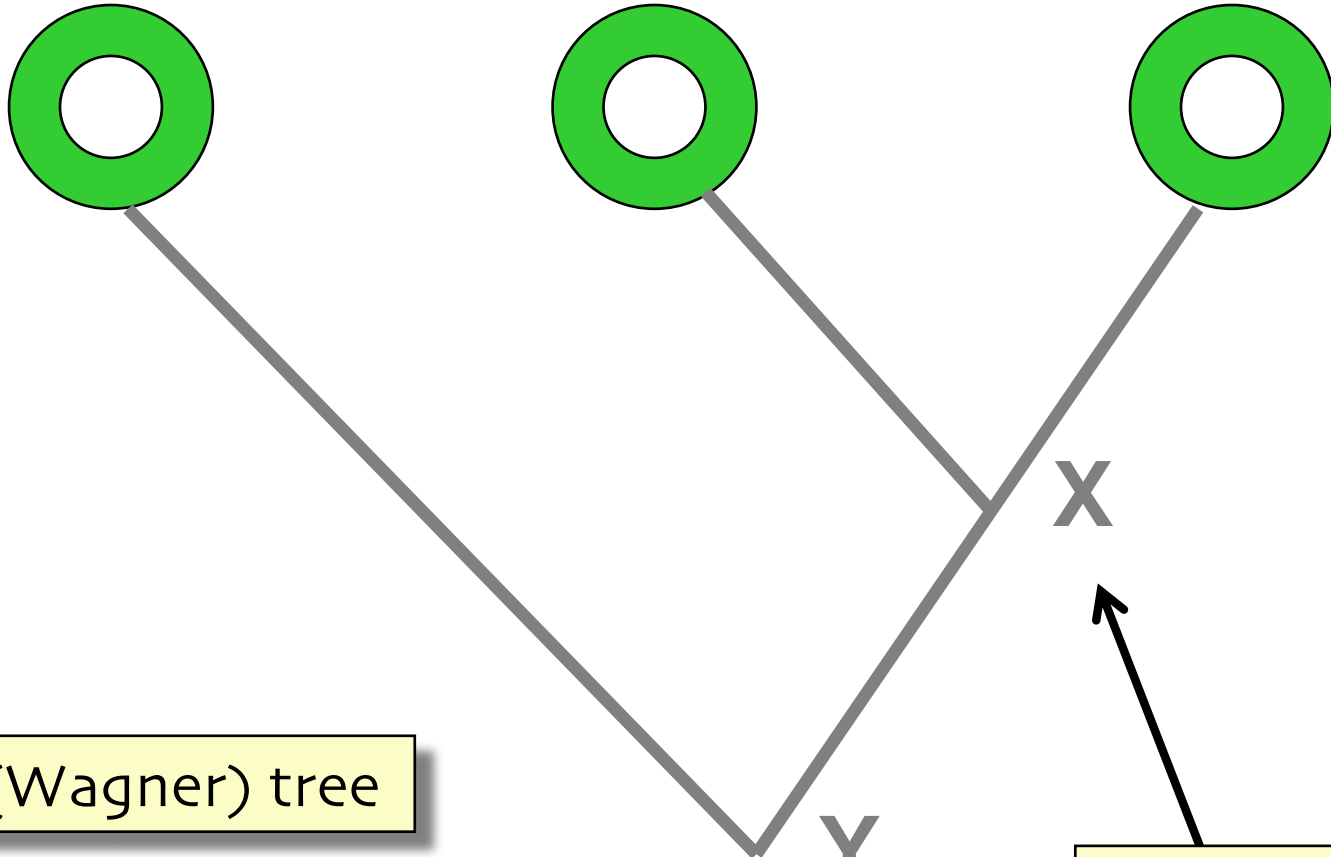
Steiner (Wagner) tree



A

B

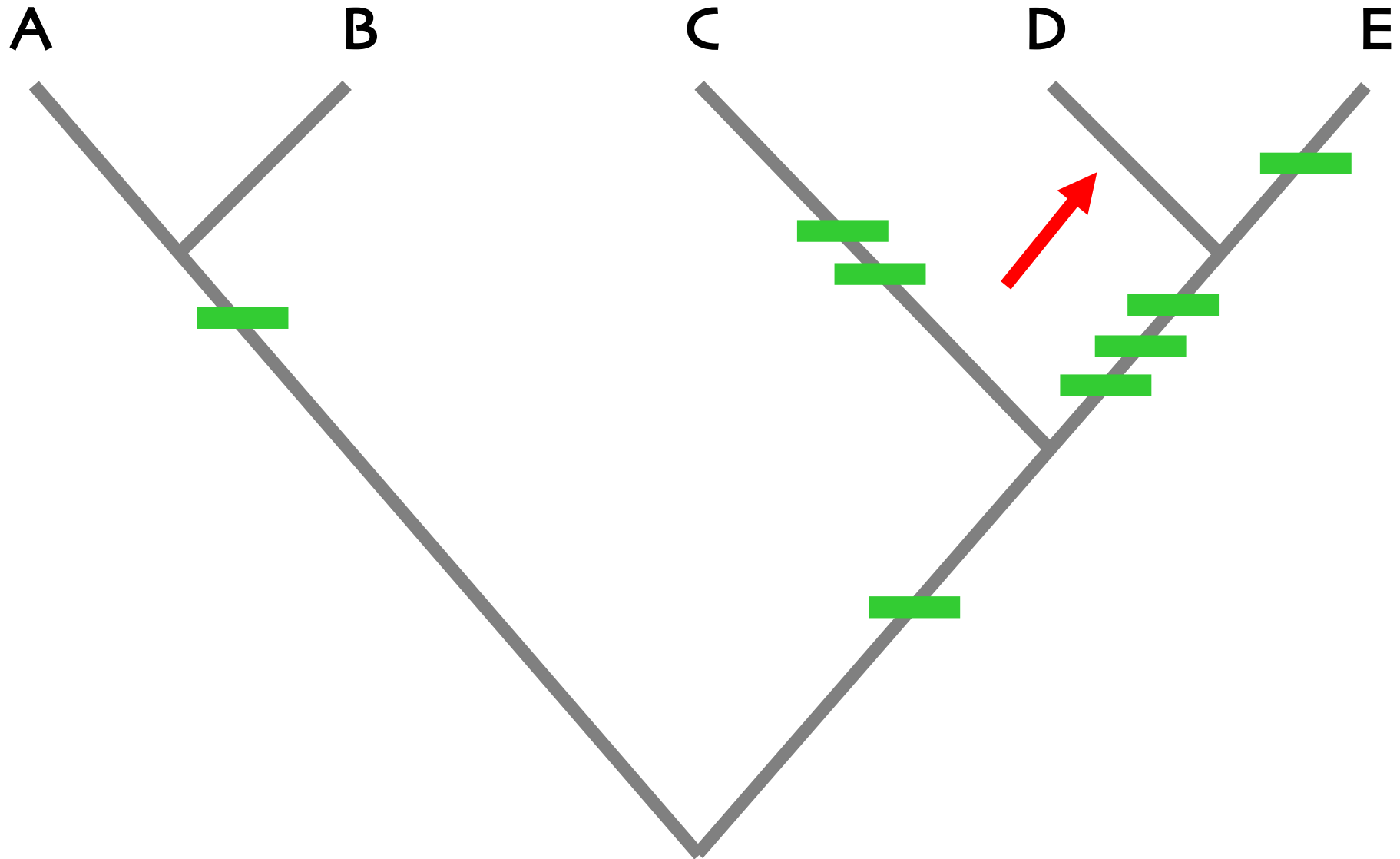
C



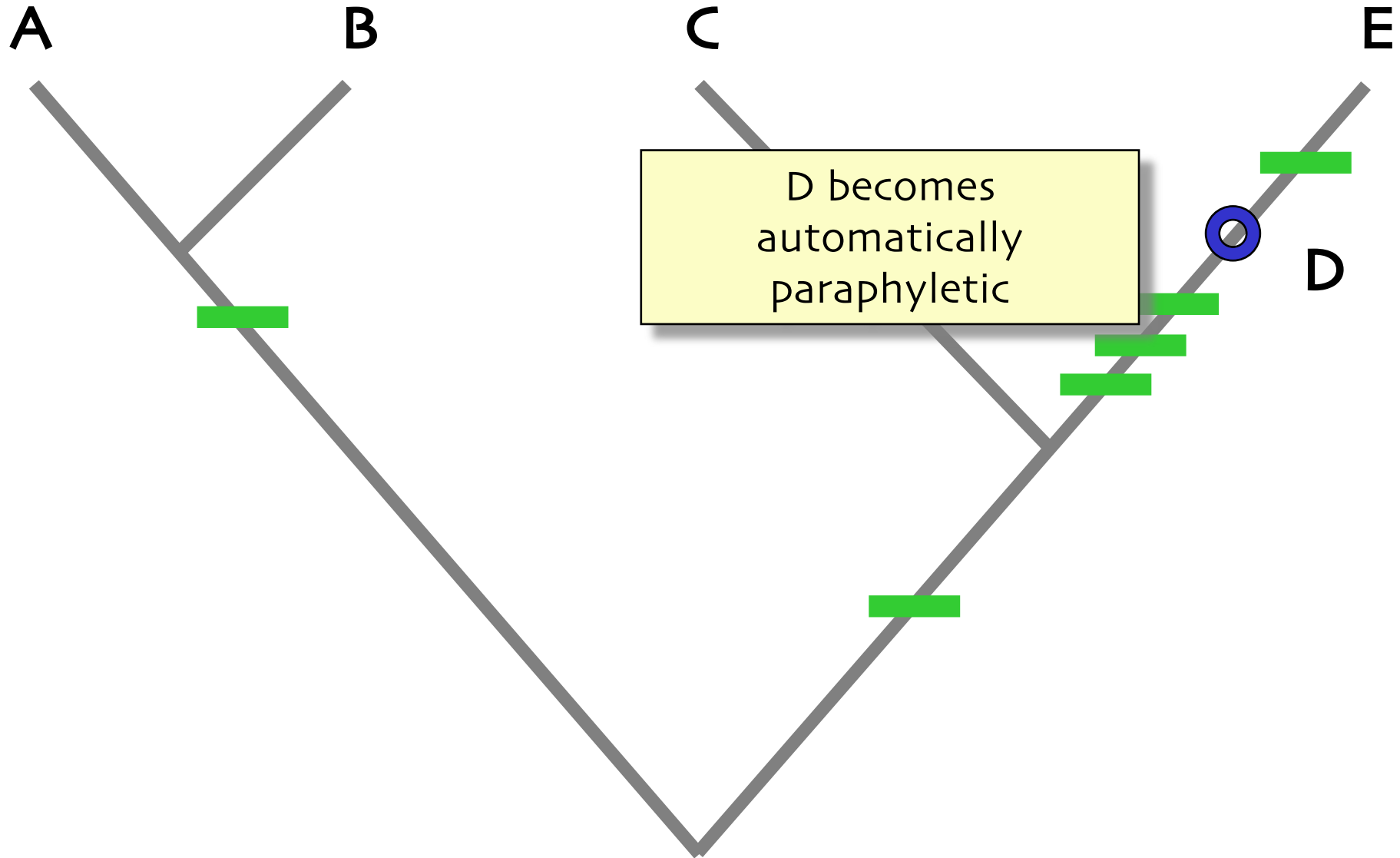
Steiner (Wagner) tree

MUCH MORE
parsimonious
than Prim tree

FOSSILS IN CLADISTIC ANALYSIS



FOSSILS IN CLADISTIC ANALYSIS



SUMMARY

fossils are **NOT** necessarily needed for phylogenetic hypotheses

they might provide valuable information

UNIQUE combinations of characters
intermediates between extant organisms

SMALL AMOUNT of information might be problematic

how missing information affects results is
unfortunately PATHOLOGICAL, i.e. CASE specific

fossils can practically never be considered as ancestors

story telling vs. science

just so –stories vs. evidence