

## 3.xi.

1. home exercise
2. character optimization
3. introduction to direct optimization
4. summary

# HOME EXERCISE

	characters
	00000000111
	123456789012
terminals	
A	010111011111
B	111000101010
C	101100100001
D	100011001000
E	100001100111

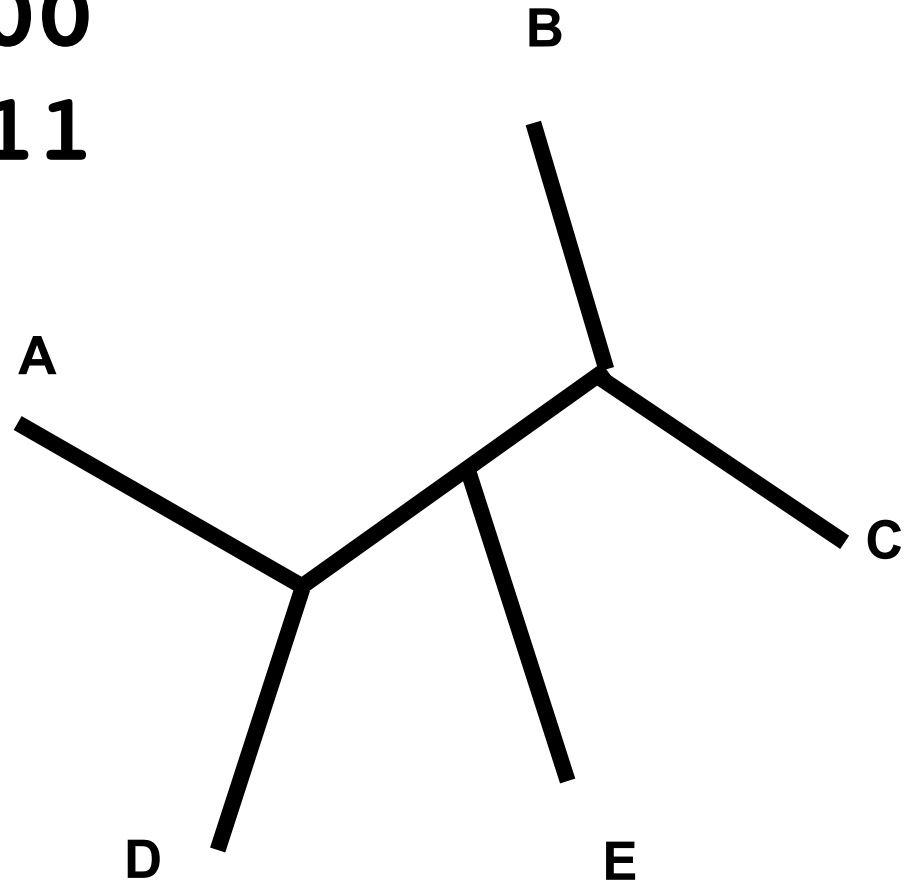
# characters

trees	0	0	0	0	0	0	0	0	0	1	1	1	$\Sigma$
	1	2	3	4	5	6	7	8	9	0	1	2	
1	1	1	2	2	2	2	2	1	2	2	1	2	20
2	1	1	2	2	2	2	2	1	2	2	2	2	21
3	1	1	2	2	2	2	2	1	1	2	2	2	20
4	1	2	2	1	2	2	2	1	2	2	2	2	21
5	1	2	2	1	2	2	2	1	2	2	2	1	20
6	1	2	2	1	2	2	2	1	2	2	2	2	21
7	1	2	2	2	2	2	2	1	2	1	1	2	20
8	1	2	1	2	2	1	2	1	2	1	2	2	19
9	1	2	2	2	2	2	2	1	2	1	2	1	20
10	1	2	2	2	2	2	2	1	1	2	2	1	20
11	1	2	2	2	2	2	2	1	2	2	1	2	21
12	1	2	1	2	2	1	2	1	2	2	2	2	20
13	1	2	2	2	1	2	1	1	1	2	2	2	19
14	1	2	2	2	1	2	1	1	2	2	2	2	20
15	1	2	1	2	1	1	1	1	2	2	2	2	18

<b>A</b>	<b>010111011111</b>
<b>B</b>	<b>111000101010</b>
<b>C</b>	<b>101100100001</b>
<b>D</b>	<b>100011001000</b>
<b>E</b>	<b>100001100111</b>

(most) parsimonious, i.e.  
shortest tree 18 (16+2)  
evolutionary changes

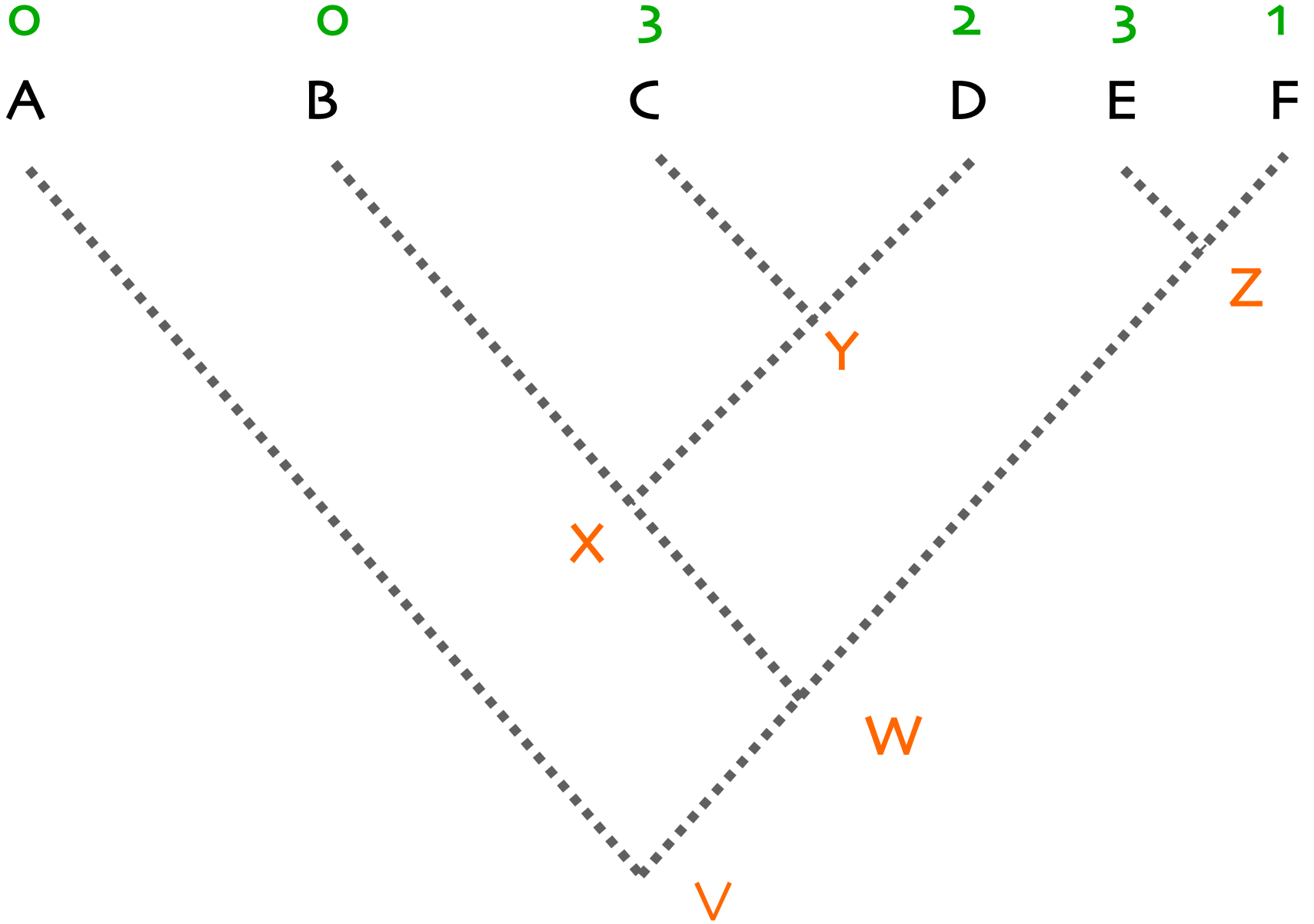
e.g. character 9  
2 evolutionary changes

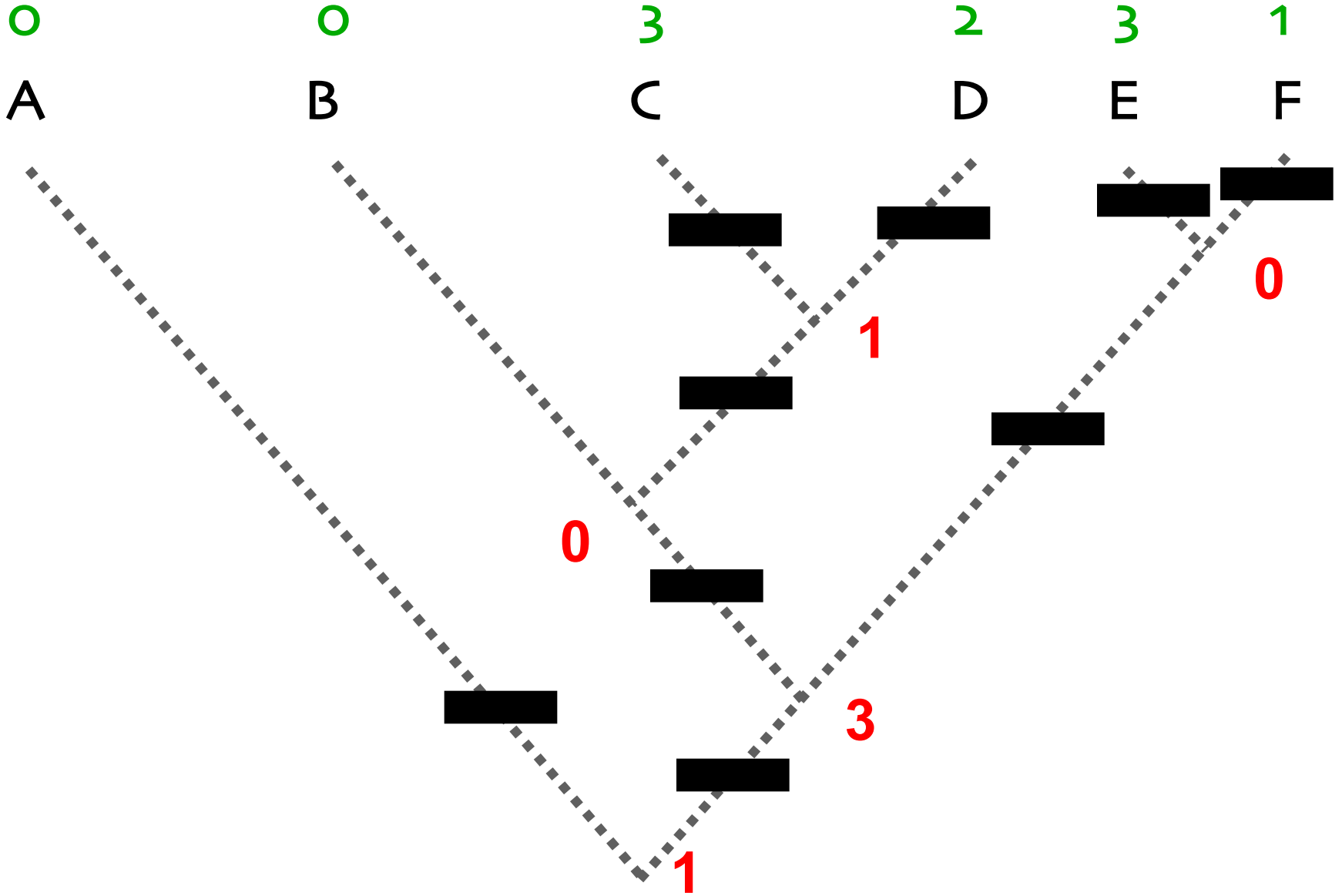


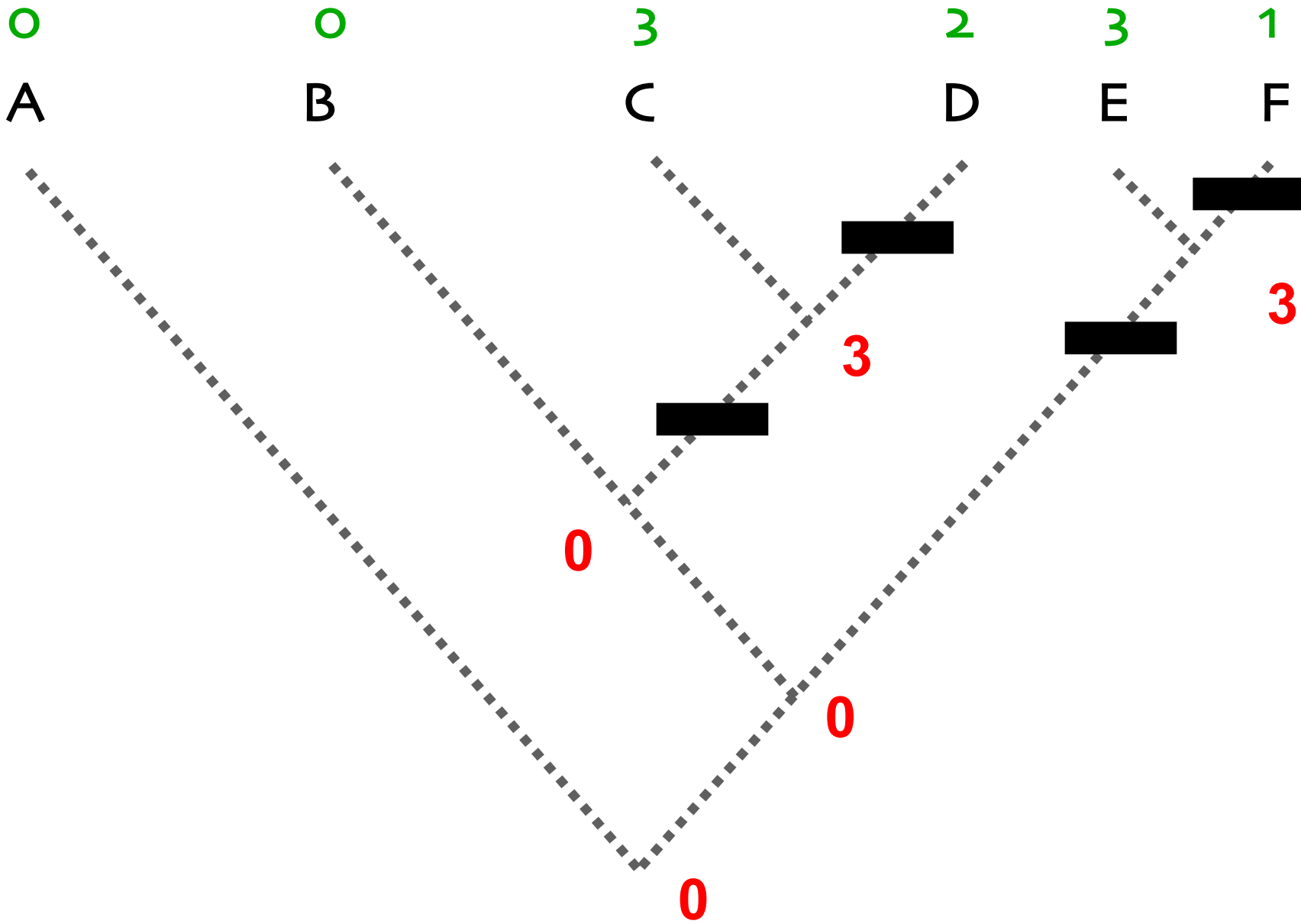
# OPTIMIZATION

HTU, Hypothetical Taxonomic Unit

hypotheses of character states for internal nodes  
(HTU) of tree









# OPTIMIZATION

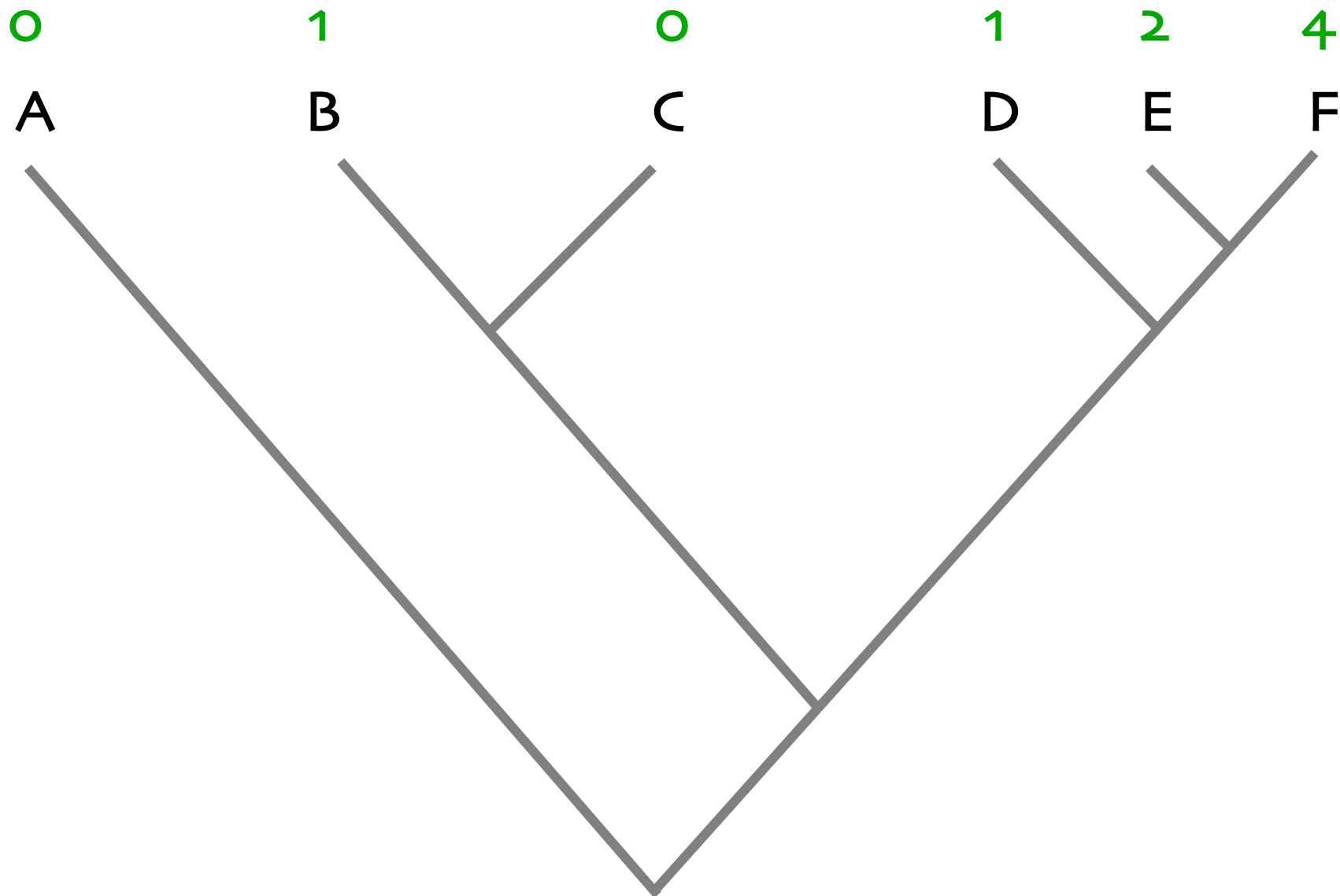
Farris, J.S. 1970. Methods for computing Wagner trees.

*Systematic Zoology* 19: 83-92.

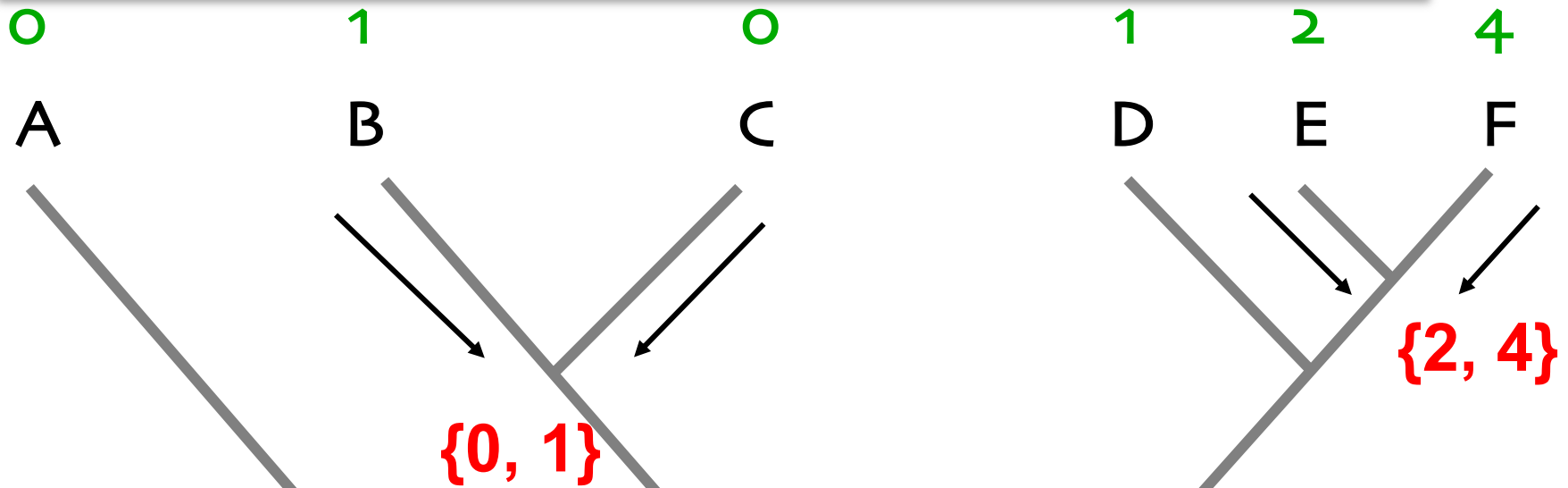
Fitch, W.M. 1971. Toward defining the course of evolution : minimal change for a specific tree topology.

*Systematic Zoology* 20: 406-416.

1 character with 4 ch. states coded with 0, 1, 2 & 4



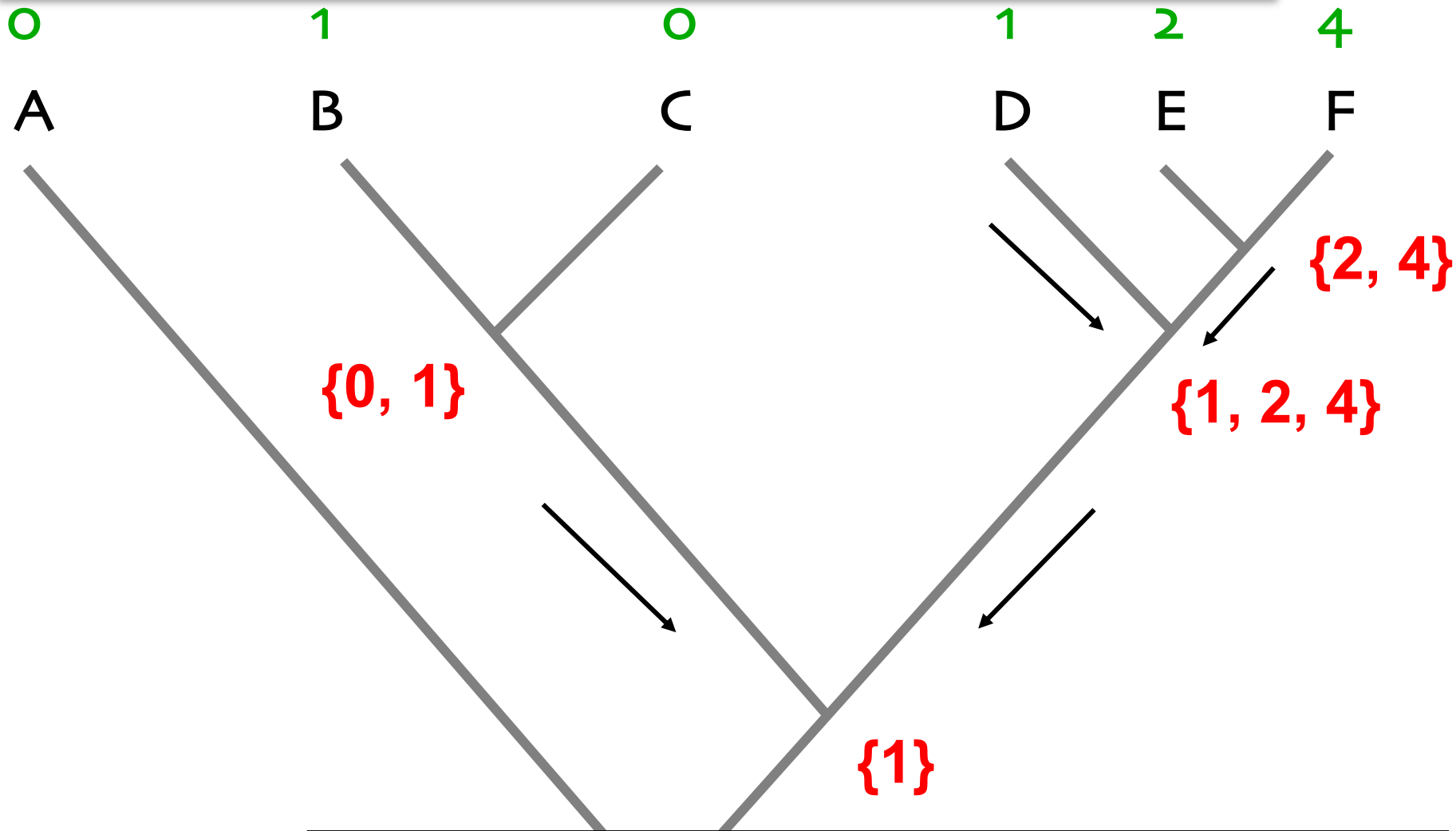
RULE 2: if terminals do not share ch. state (intersection,  $\cap = \emptyset$ )  
their (union,  $\cup$ ) is marked for their ancestor



RULE 1: if terminals share character state this will  
be marked also for their ancestor (intersection,  $\cap$ )

DOWNWARD  
PASS = postorder

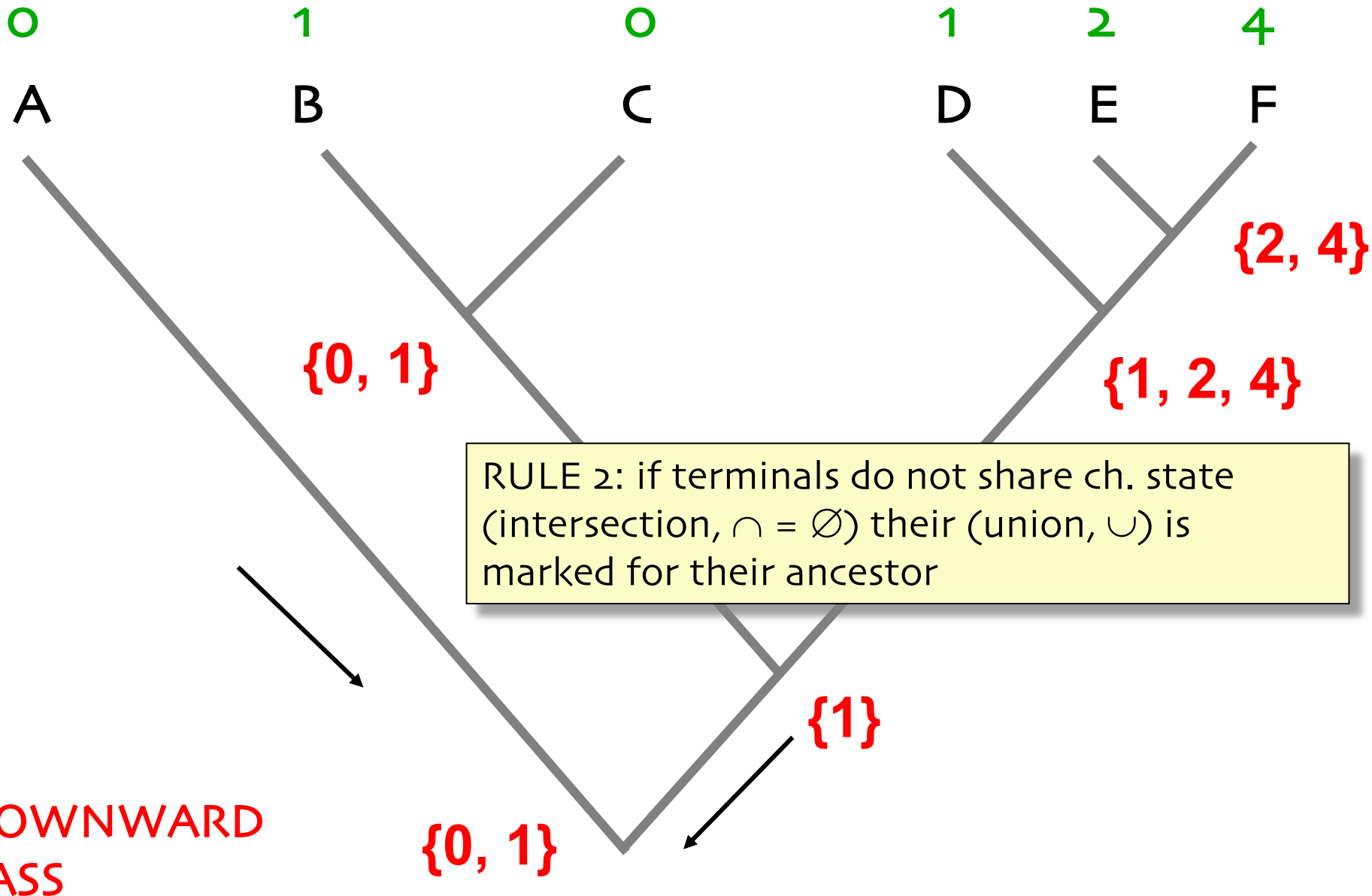
RULE 2: if terminals do not share ch. state (intersection,  $\cap = \emptyset$ )  
their (union,  $\cup$ ) is marked for their ancestor



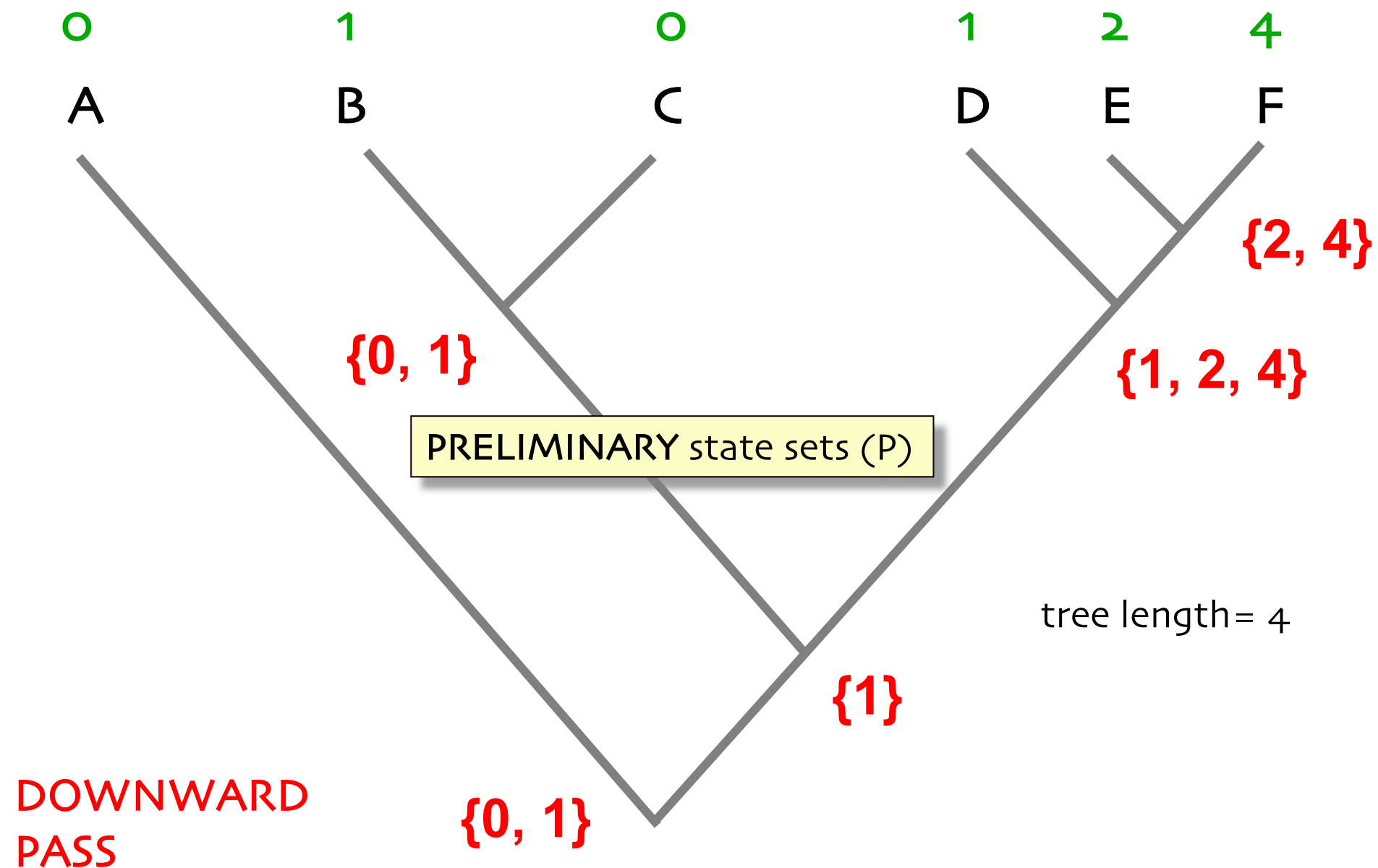
**DOWNWARD  
PASS**

RULE 1: if terminals share character state this will  
be marked also for their ancestor (intersection,  $\cap$ )

RULE 1: if terminals share character state this will be marked also for their ancestor (intersection,  $\cap$ )



ATTENTION! LENGTH of diagram, number of ch. state changes, calculated already at this stage. Unions ( $\cup$ ) add always one ch. state change.



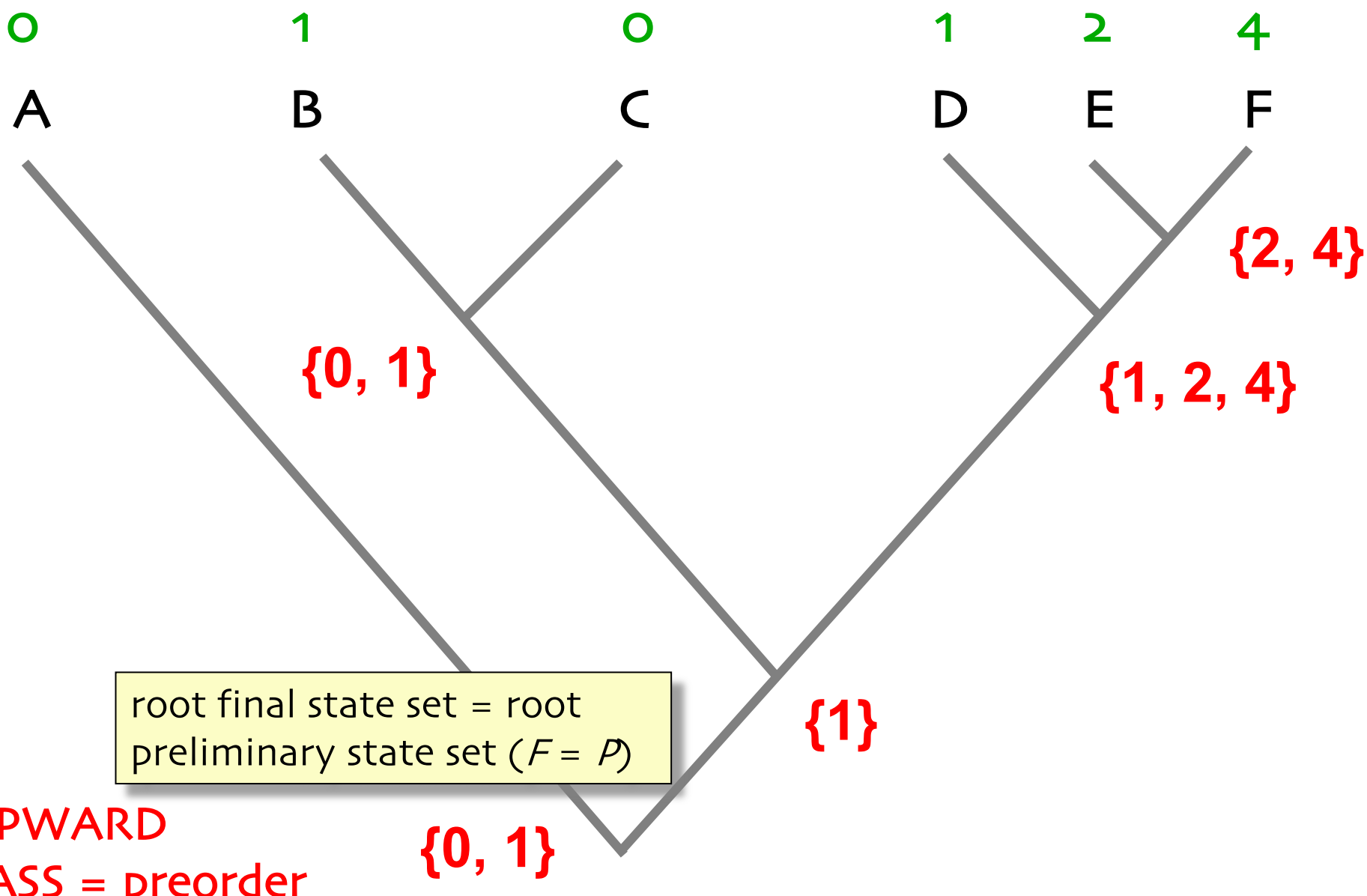
“upward pass” rules (Fitch 1971, Wheeler 2012)

- root final state set = root preliminary state set ( $F = P$ )

RULE 1. If the intersection of the preliminary state,  $P$ , of the node and its ancestor,  $A$ , is equal to  $A$ , (if  $A \cap P = A$ ) then the final state set,  $F$ , is equal to that of the ancestor ( $F = A$ ).

RULE 2. If Rule 1 does not apply and the union of final/preliminary states of the 2 descendants of the current node (*Left* and *Right*) are equal to preliminary states of the current node ( $P = L \cup R$ ), then  $F = P \cup A$ .

RULE 3. If Rule 1 and 2 do not apply the final state set is the preliminary state set, supplemented by state set that is common to the ancestor and descendants ( $F = P \cup (L \cap A) \cup (R \cap A)$ ).



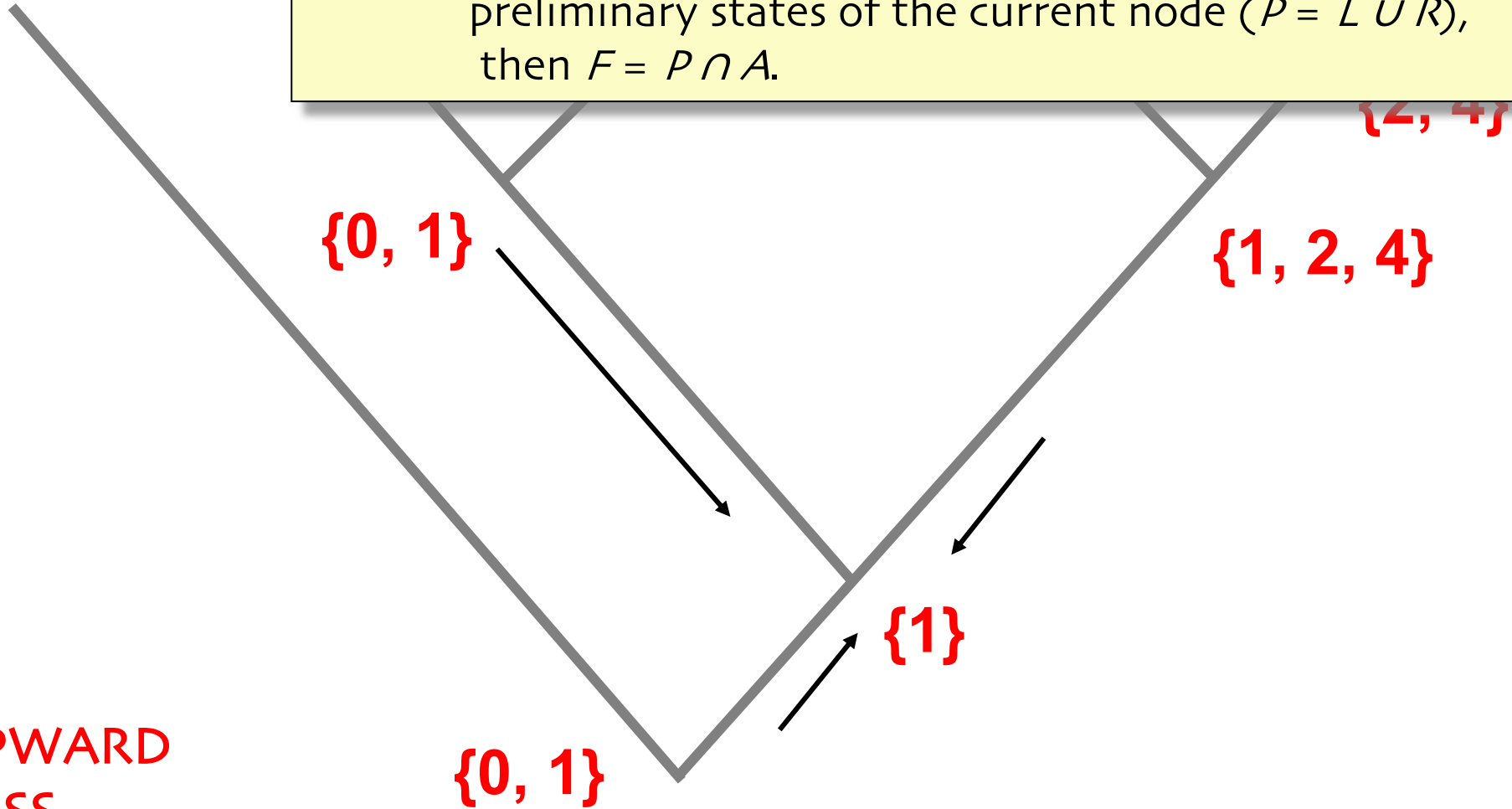


RULE 1. If the overlap of the preliminary state,  $P$ , of the node and its ancestor,  $A$ , is equal to  $A$ , (if  $A \cap P = A$ ) then the final state set,  $F$ , is equal to that of the ancestor ( $F = A$ ).

RULE 2. If Rule 1 does not apply and the union of final/preliminary states of the 2 descendants of the current node ( $Left$  and  $Right$ ) are equal to the preliminary states of the current node ( $P = L \cup R$ ), then  $F = P \cap A$ .

○

A



UPWARD  
PASS

O  
A

RULE 3. If Rules 1 and 2 do not apply the final state set is the preliminary state set, supplemented by state set that is common to the ancestor and descendants  
 $(F = P \cup (L \cap A) \cup (R \cap A))$ .

$$= 1 \cup (0,1 \cap 0,1) \cup (1,2,4 \cap 0,1)$$

$$= 1 \cup (0,1) \cup (1)$$

$$= 1 \cup 0$$

$$0,1$$

{2, 4}

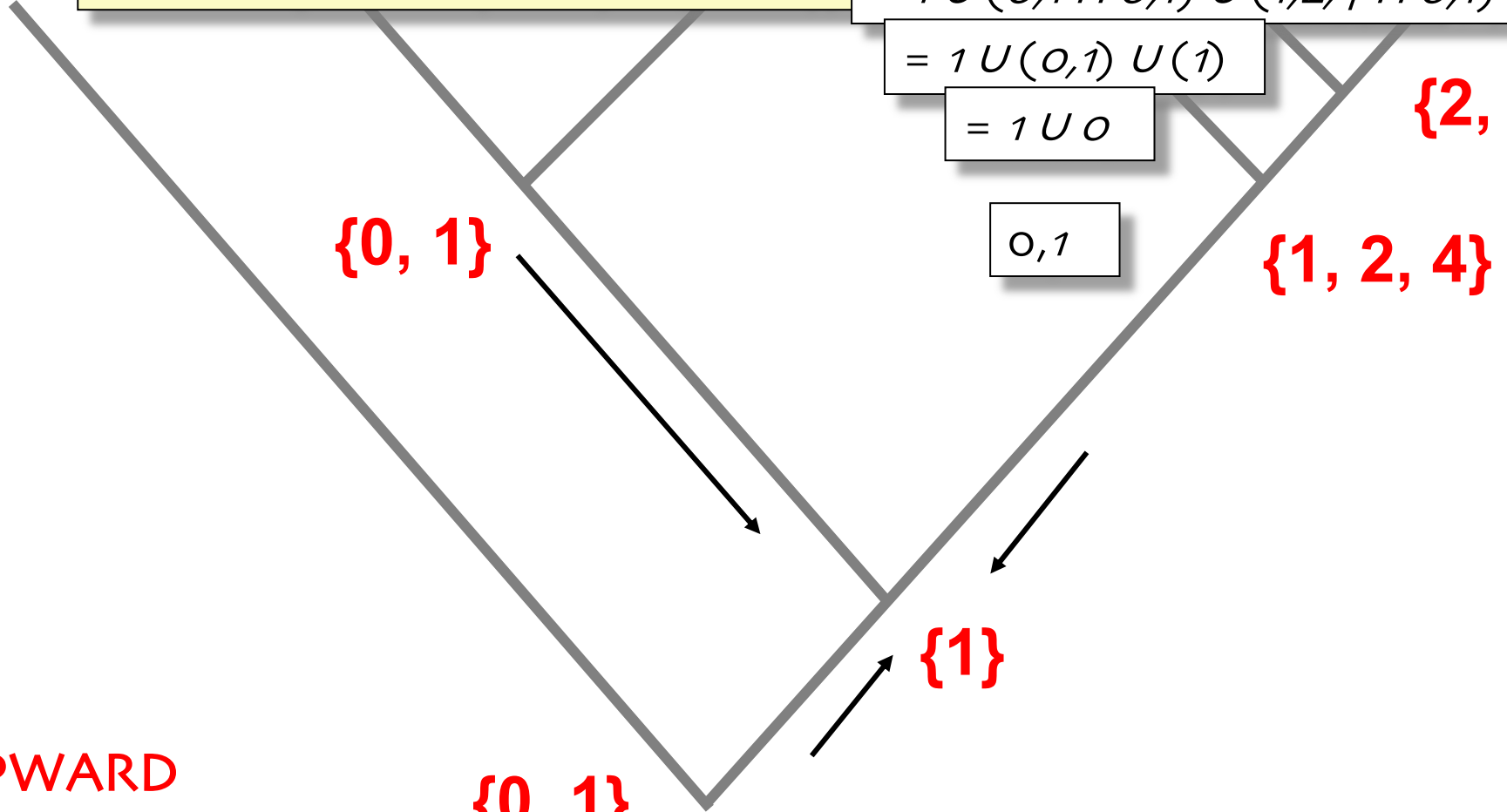
{0, 1}

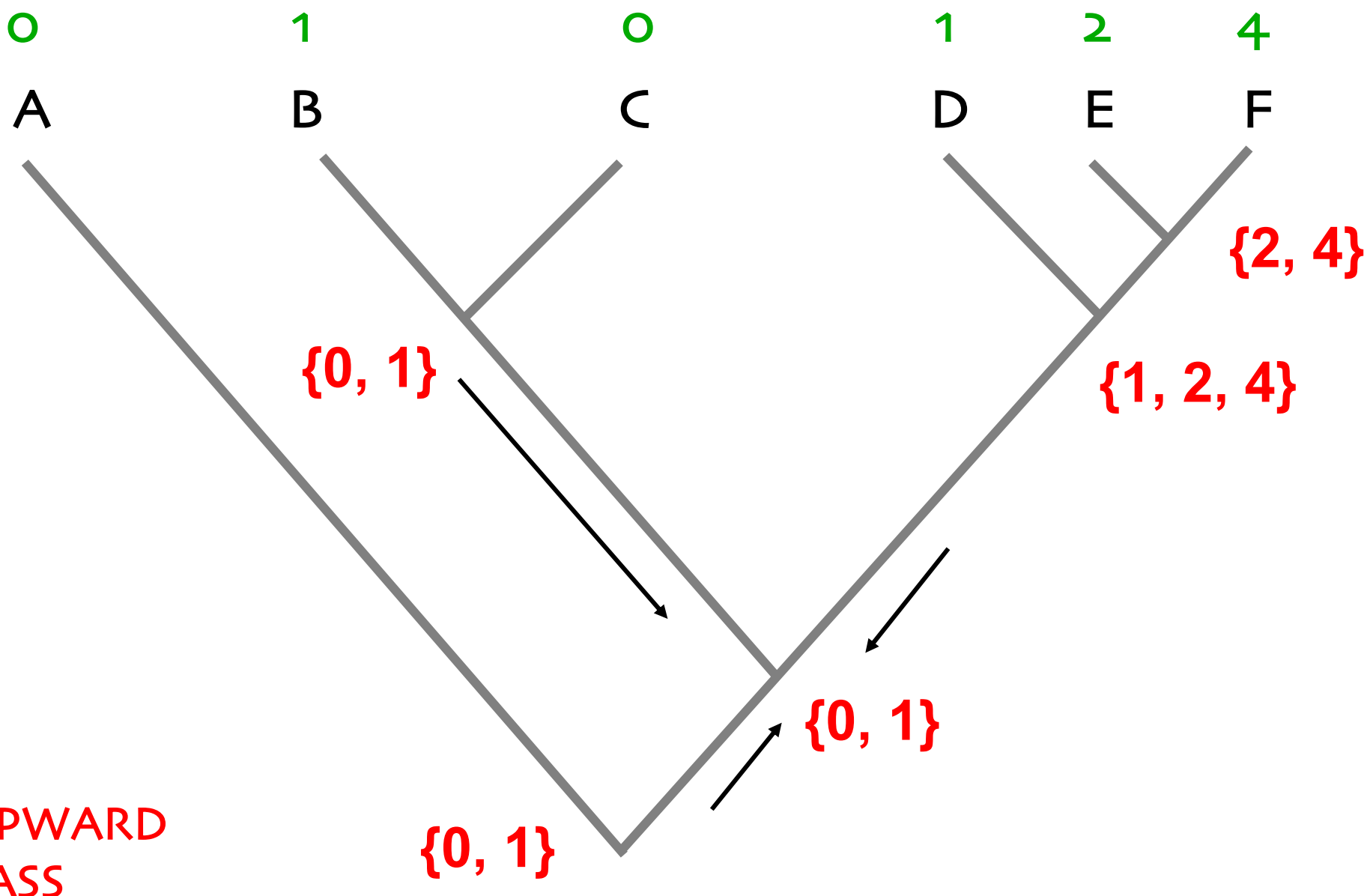
{1, 2, 4}

{1}

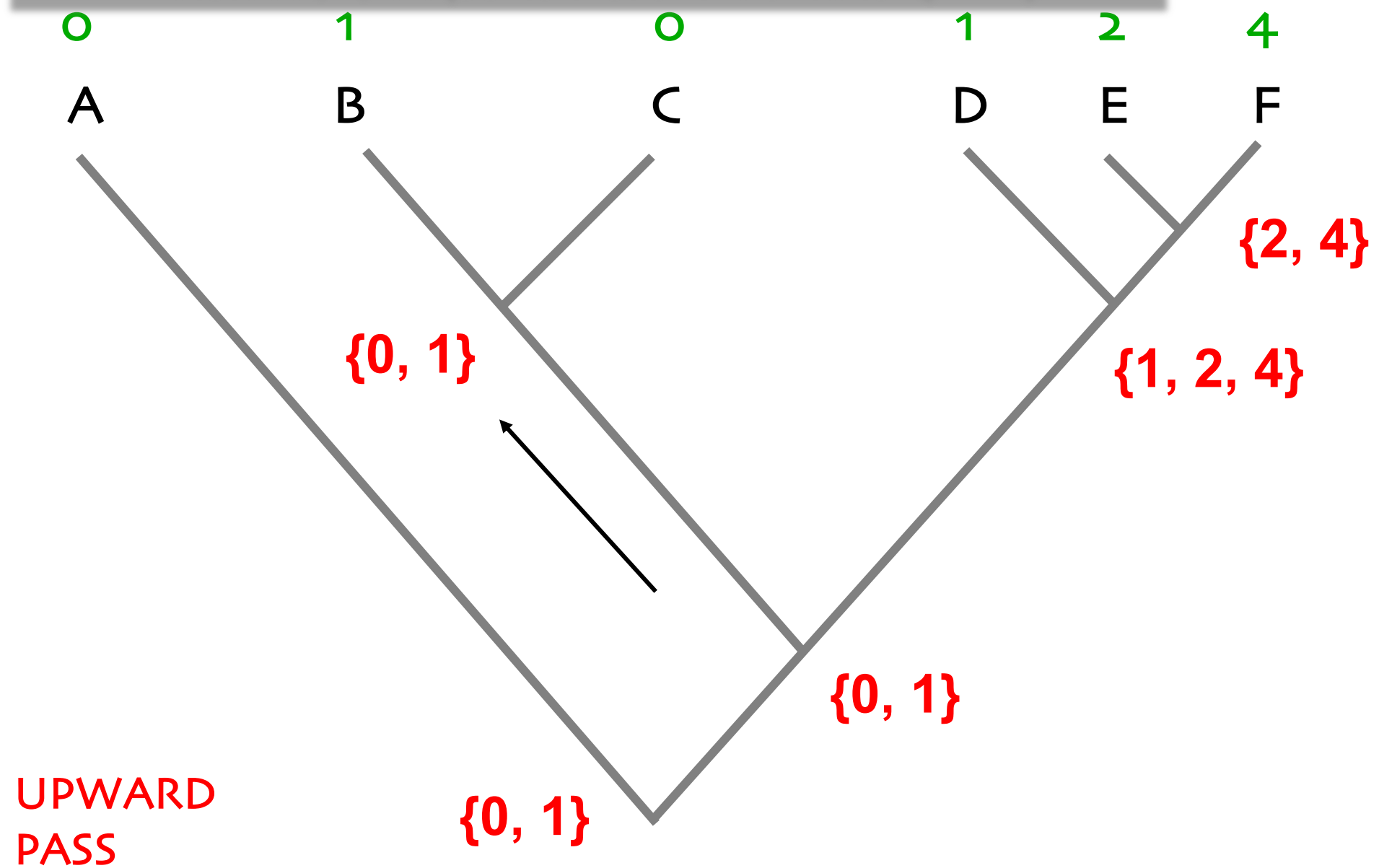
{0, 1}

UPWARD  
PASS

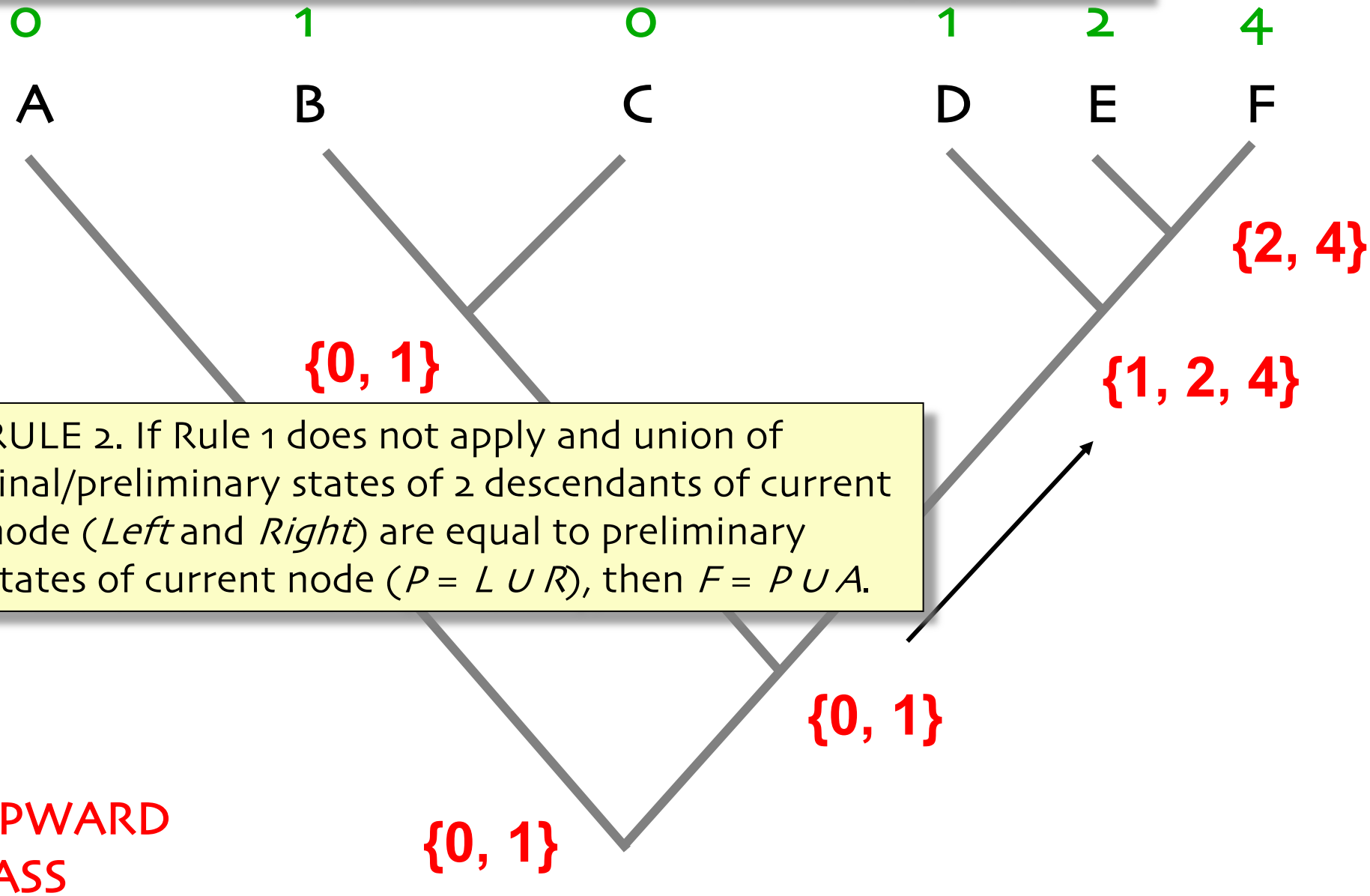




RULE 1. If the overlap of the preliminary state,  $P$ , of the node and its ancestor,  $A$ , is equal to  $A$ , (if  $A \cap P = A$ ) then the final state set,  $F$ , is equal to that of the ancestor ( $F = A$ ).

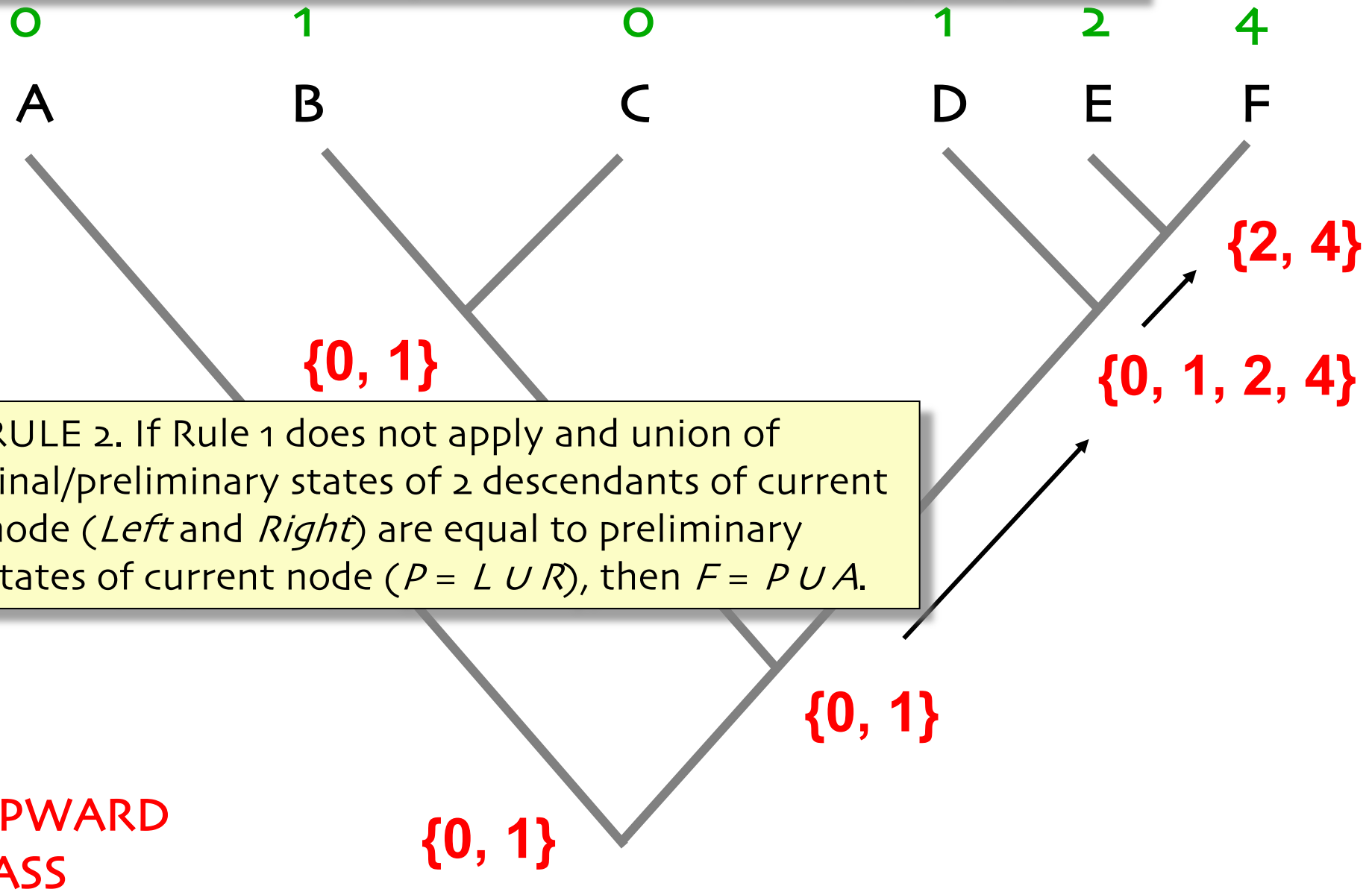


RULE 1. If the overlap of the preliminary state,  $P$ , of the node and its ancestor,  $A$ , is equal to  $A$ , (if  $A \cap P = A$ ) then the final state set,  $F$ , is equal to that of the ancestor ( $F = A$ ).



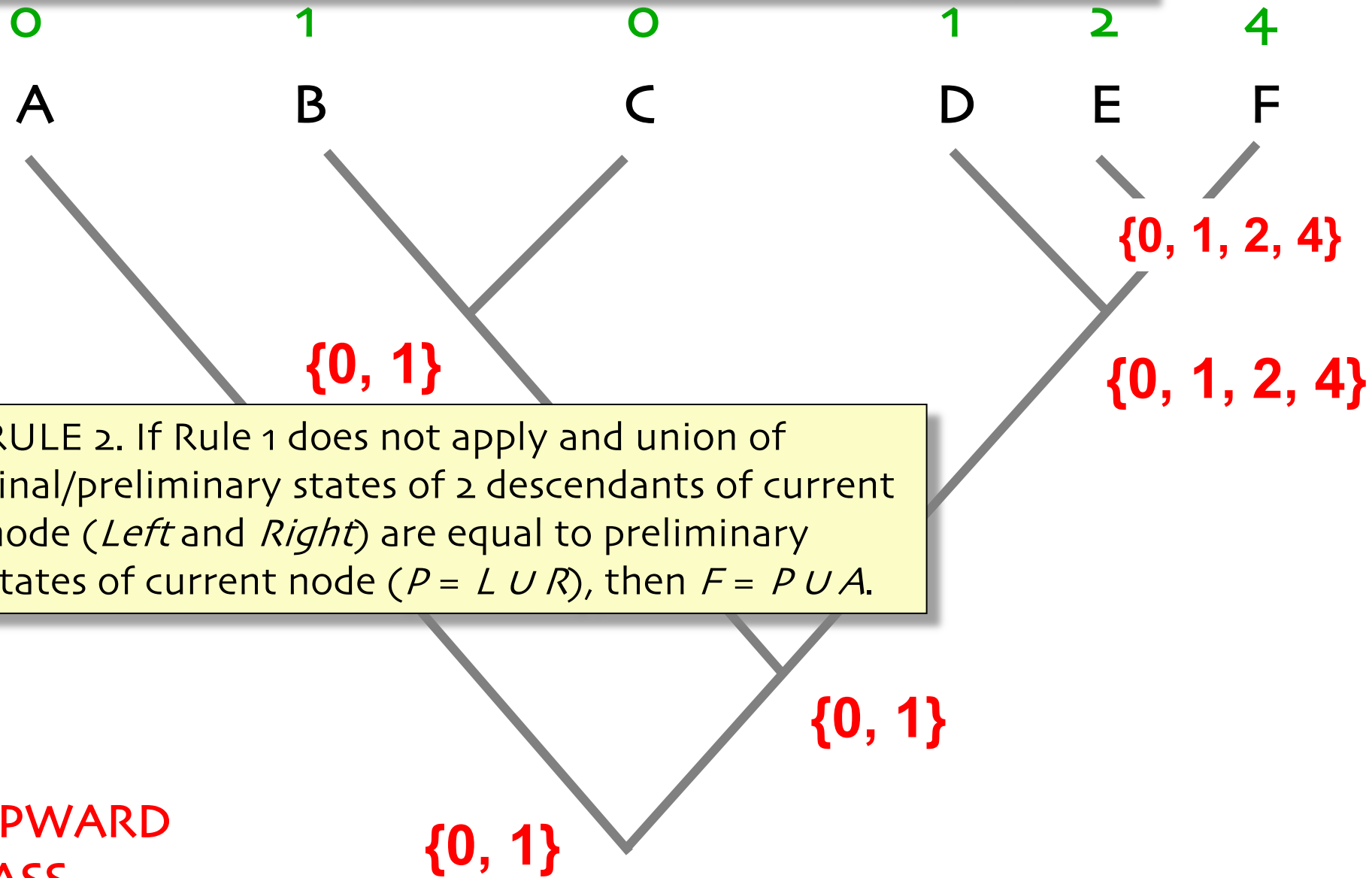
RULE 2. If Rule 1 does not apply and union of final/preliminary states of 2 descendants of current node (*Left* and *Right*) are equal to preliminary states of current node ( $P = L \cup R$ ), then  $F = P \cup A$ .

RULE 1. If the overlap of the preliminary state,  $P$ , of the node and its ancestor,  $A$ , is equal to  $A$ , (if  $A \cap P = A$ ) then the final state set,  $F$ , is equal to that of the ancestor ( $F = A$ ).



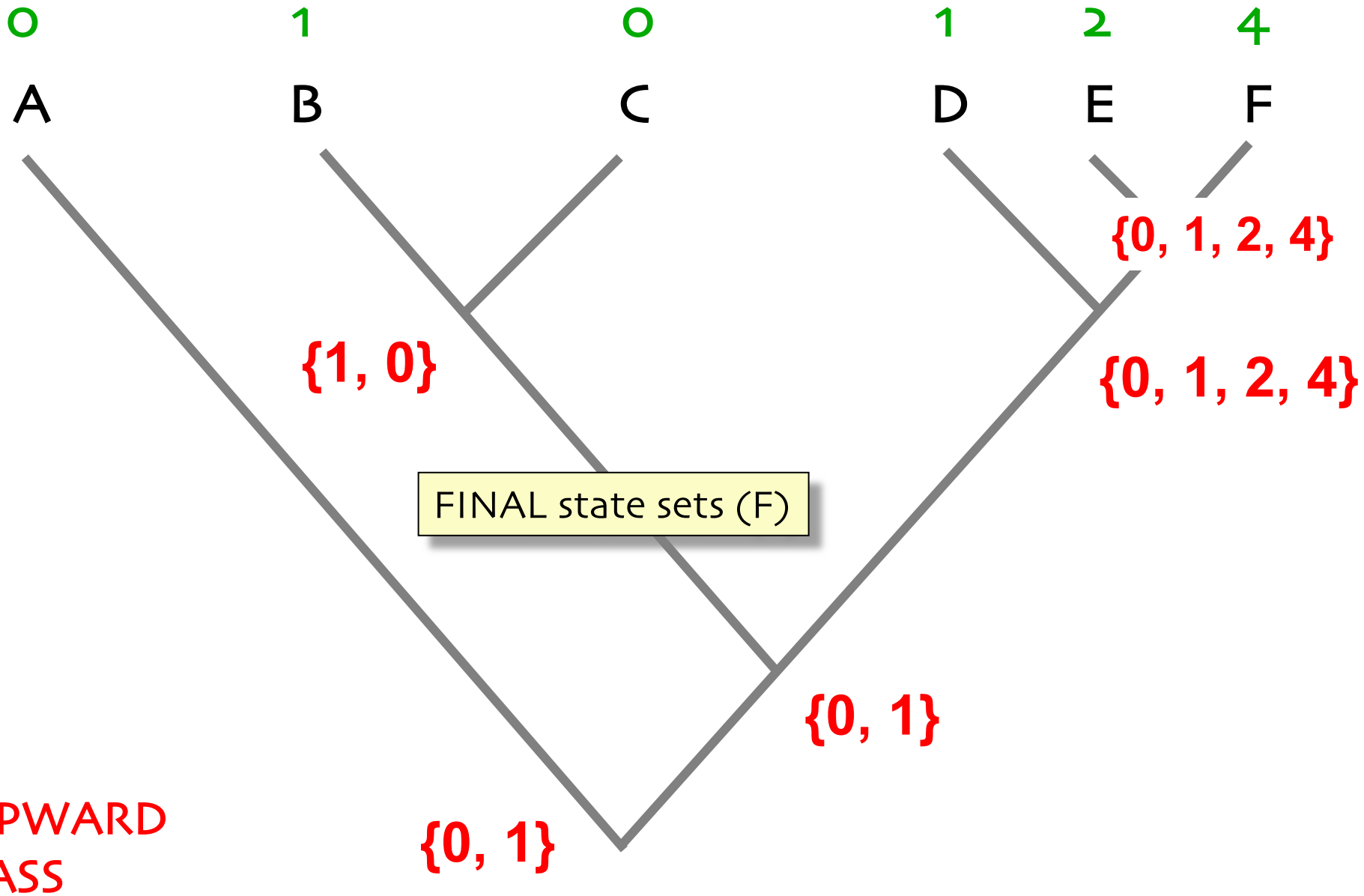
RULE 2. If Rule 1 does not apply and union of final/preliminary states of 2 descendants of current node (*Left* and *Right*) are equal to preliminary states of current node ( $P = L \cup R$ ), then  $F = P \cup A$ .

RULE 1. If the overlap of the preliminary state,  $P$ , of the node and its ancestor,  $A$ , is equal to  $A$ , (if  $A \cap P = A$ ) then the final state set,  $F$ , is equal to that of the ancestor ( $F = A$ ).

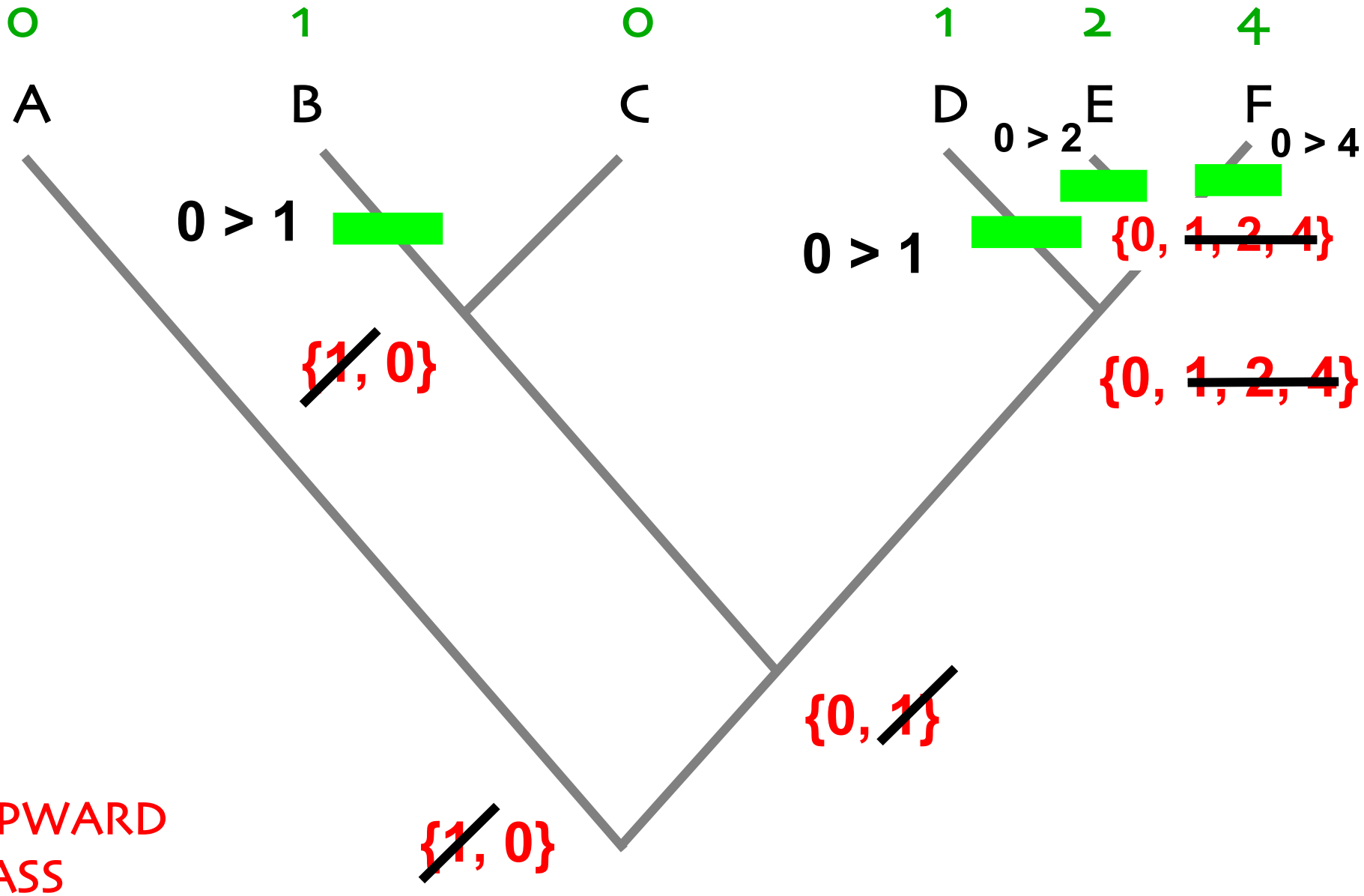


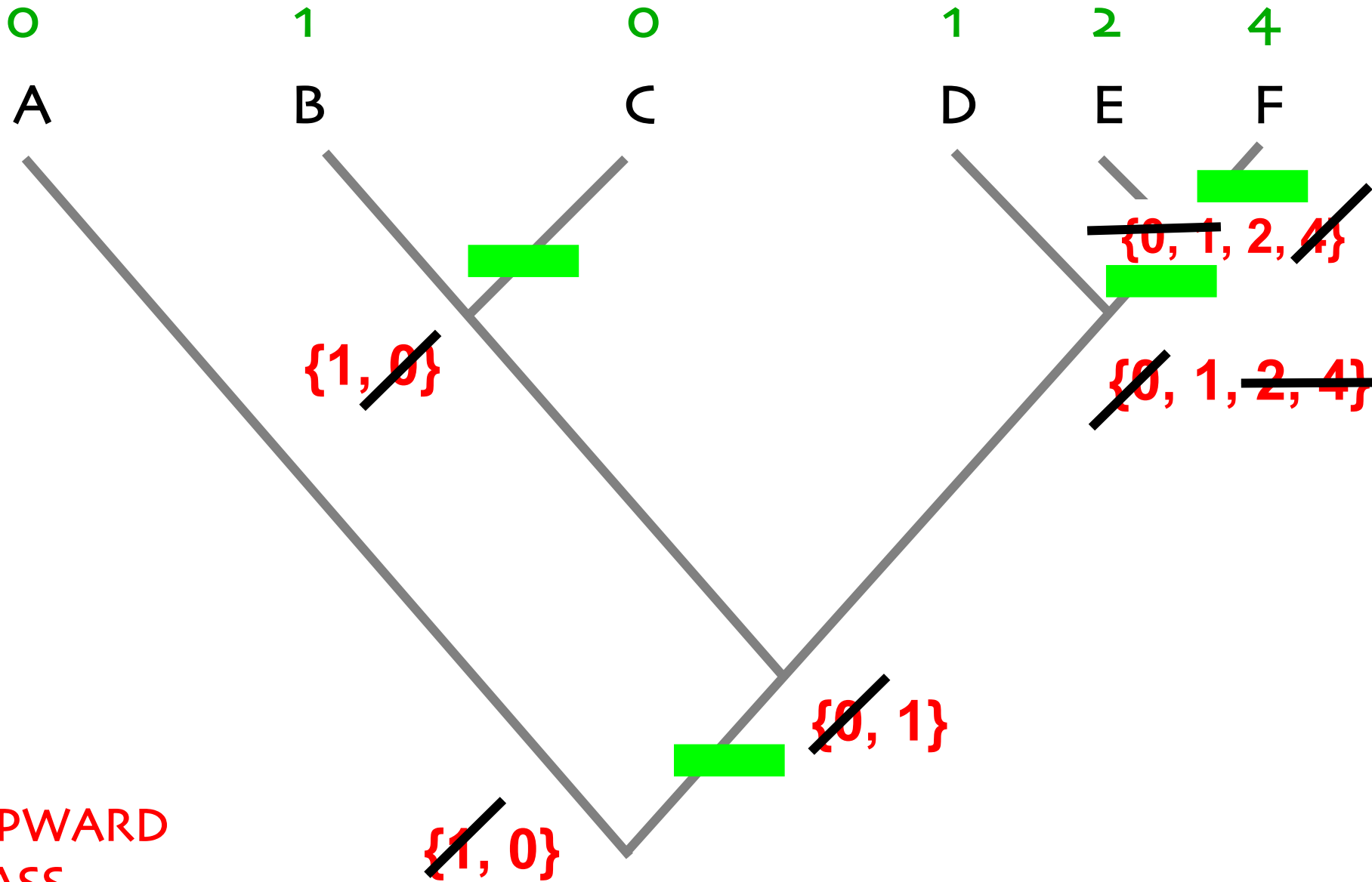
RULE 2. If Rule 1 does not apply and union of final/preliminary states of 2 descendants of current node (*Left* and *Right*) are equal to preliminary states of current node ( $P = L \cup R$ ), then  $F = P \cup A$ .

UPWARD  
PASS

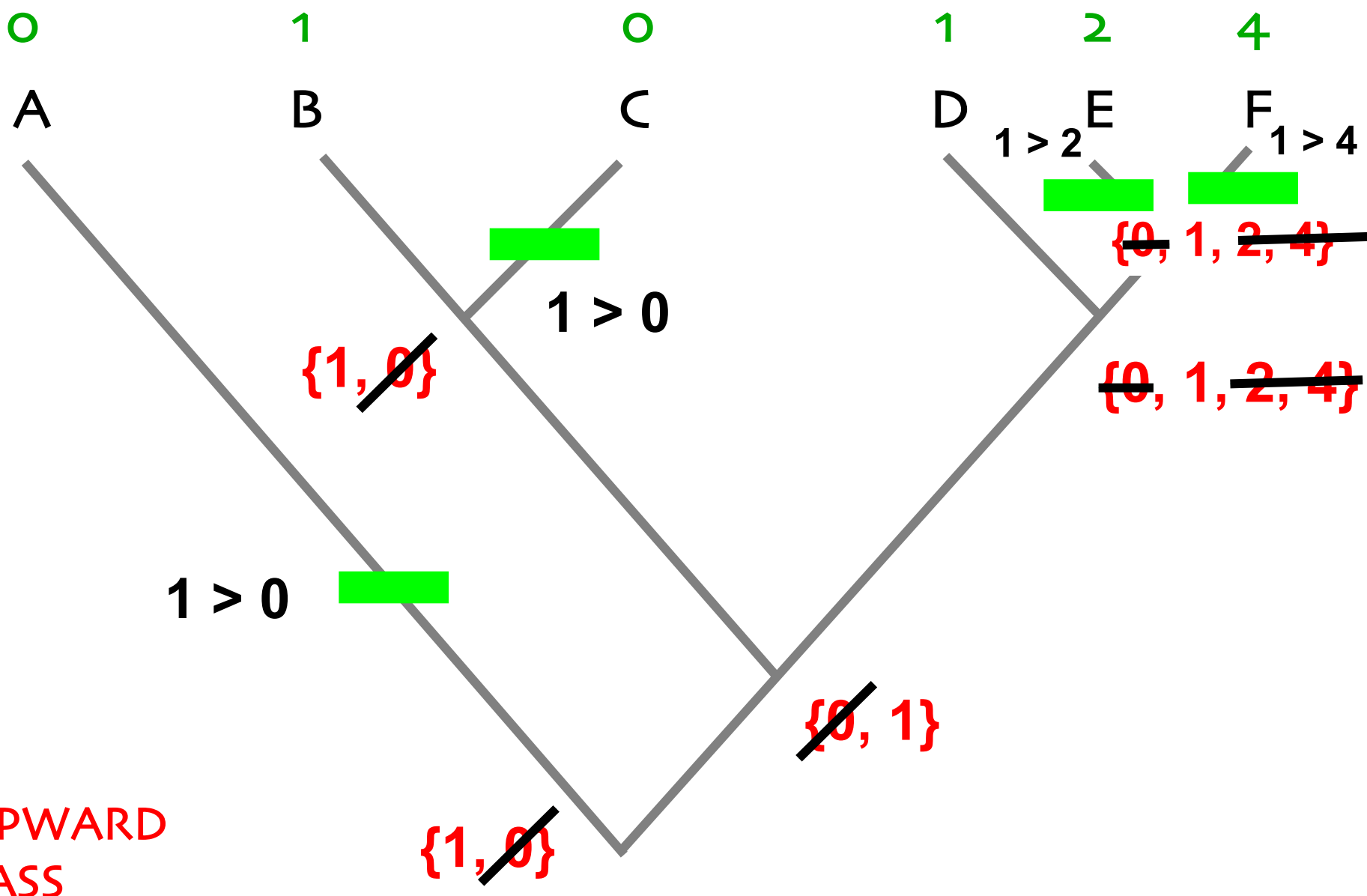


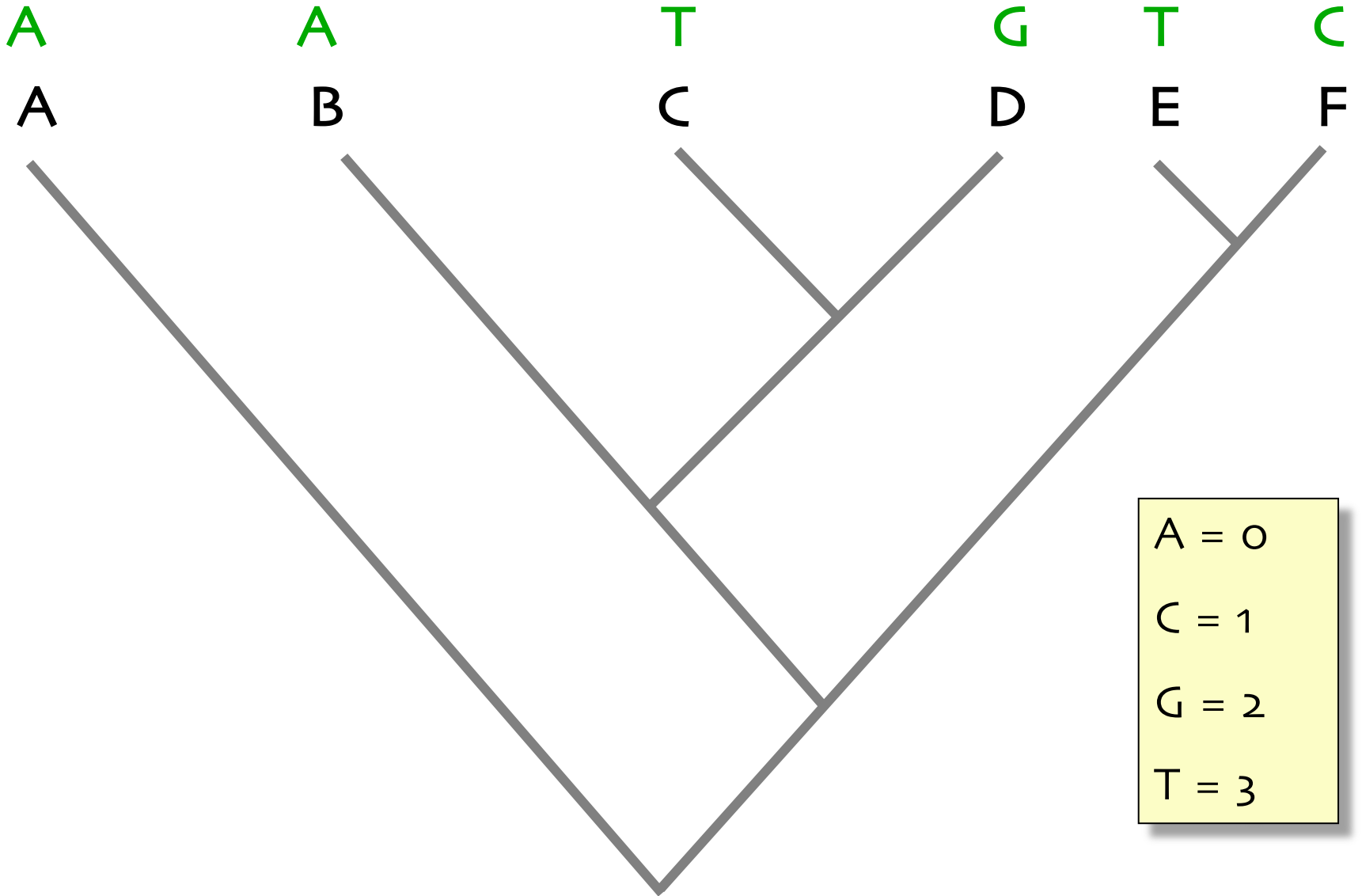




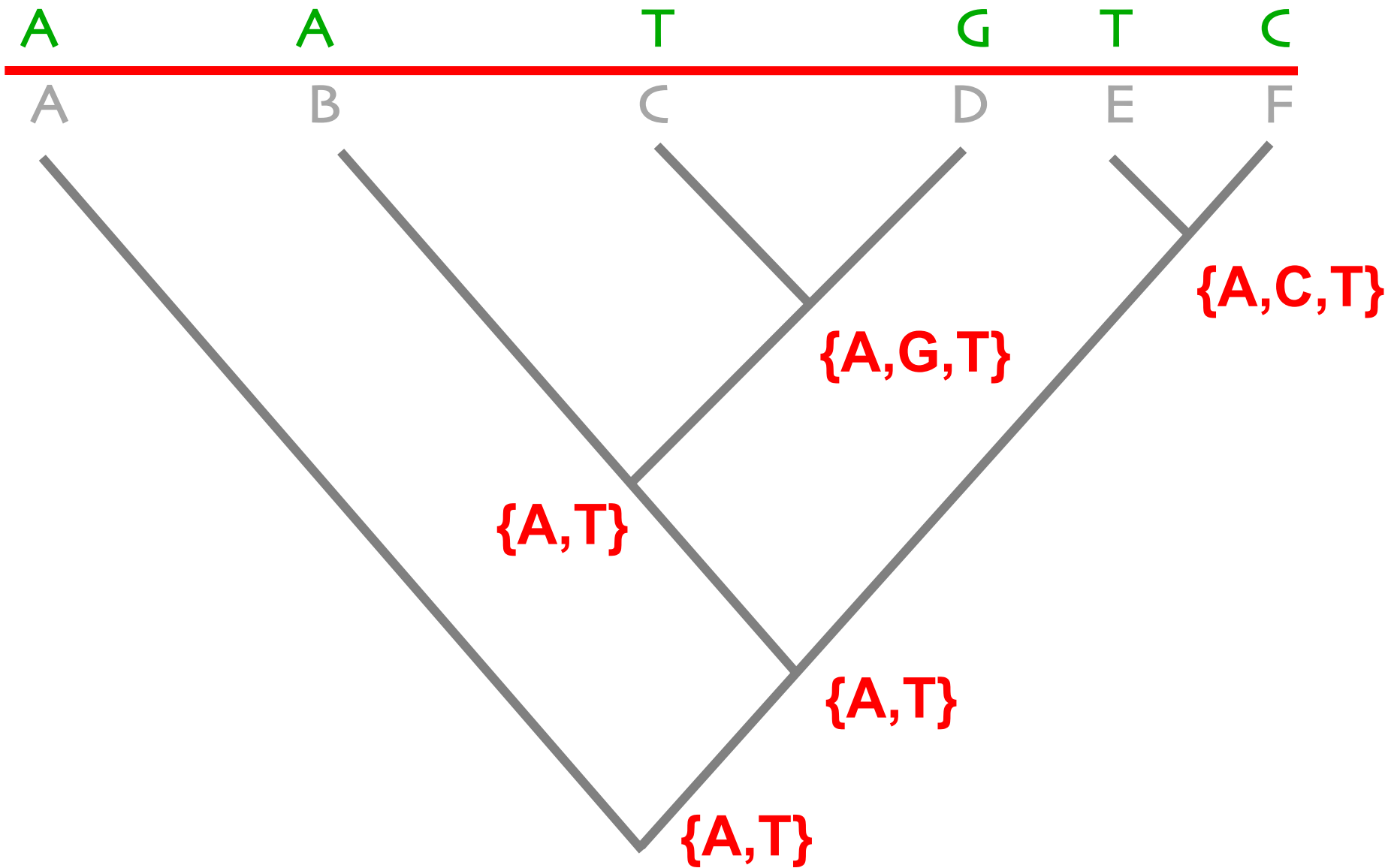


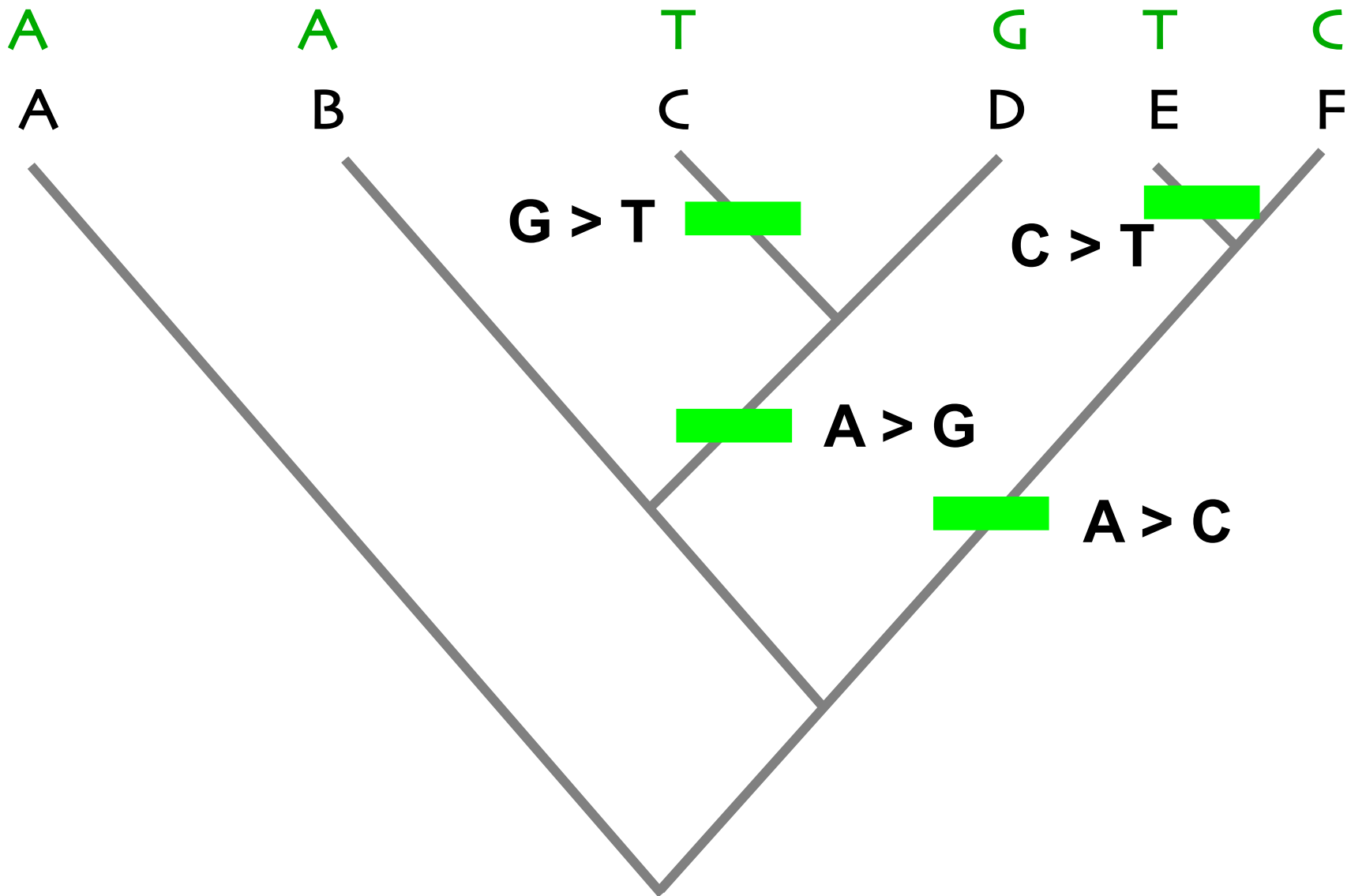
UPWARD  
PASS

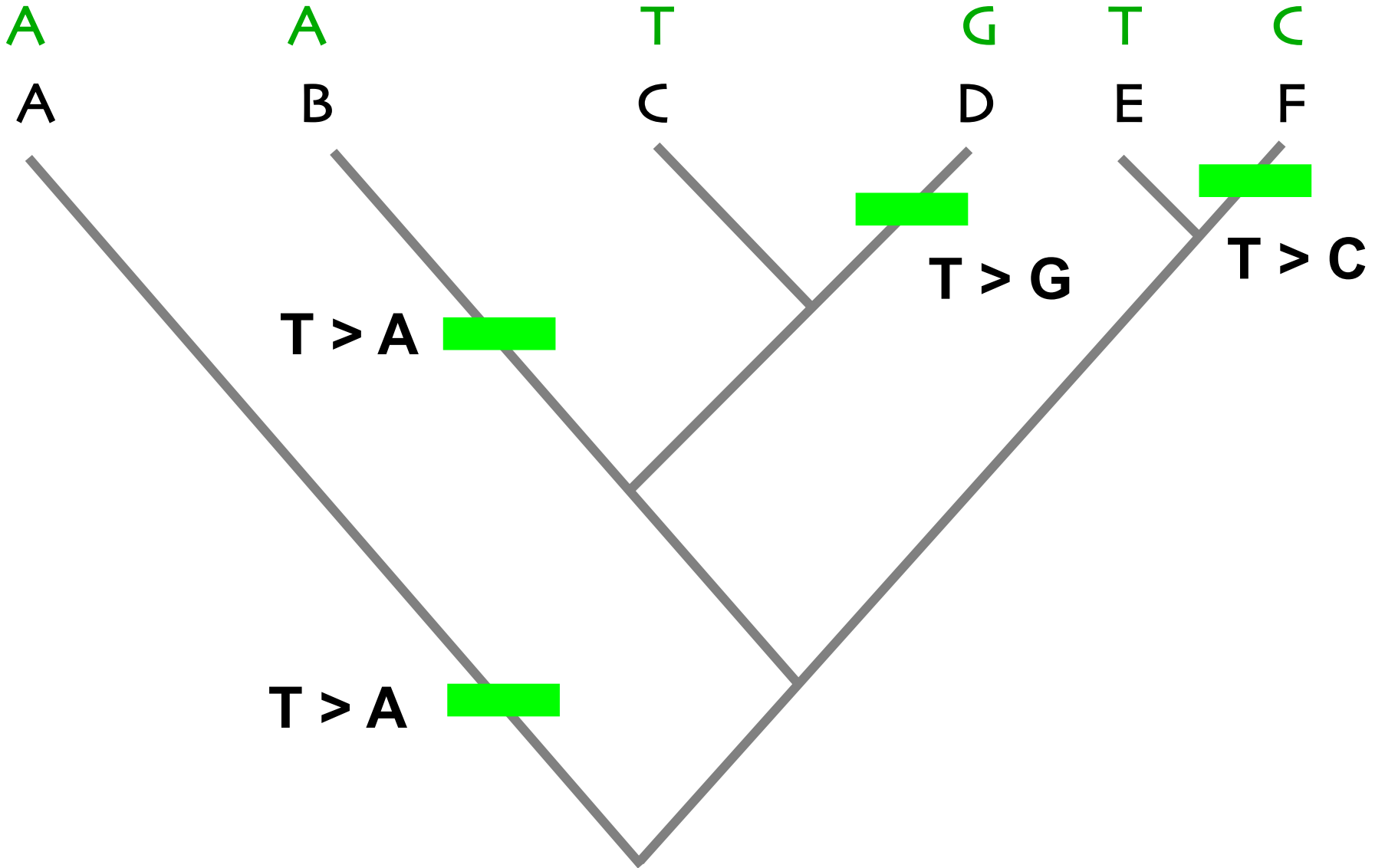


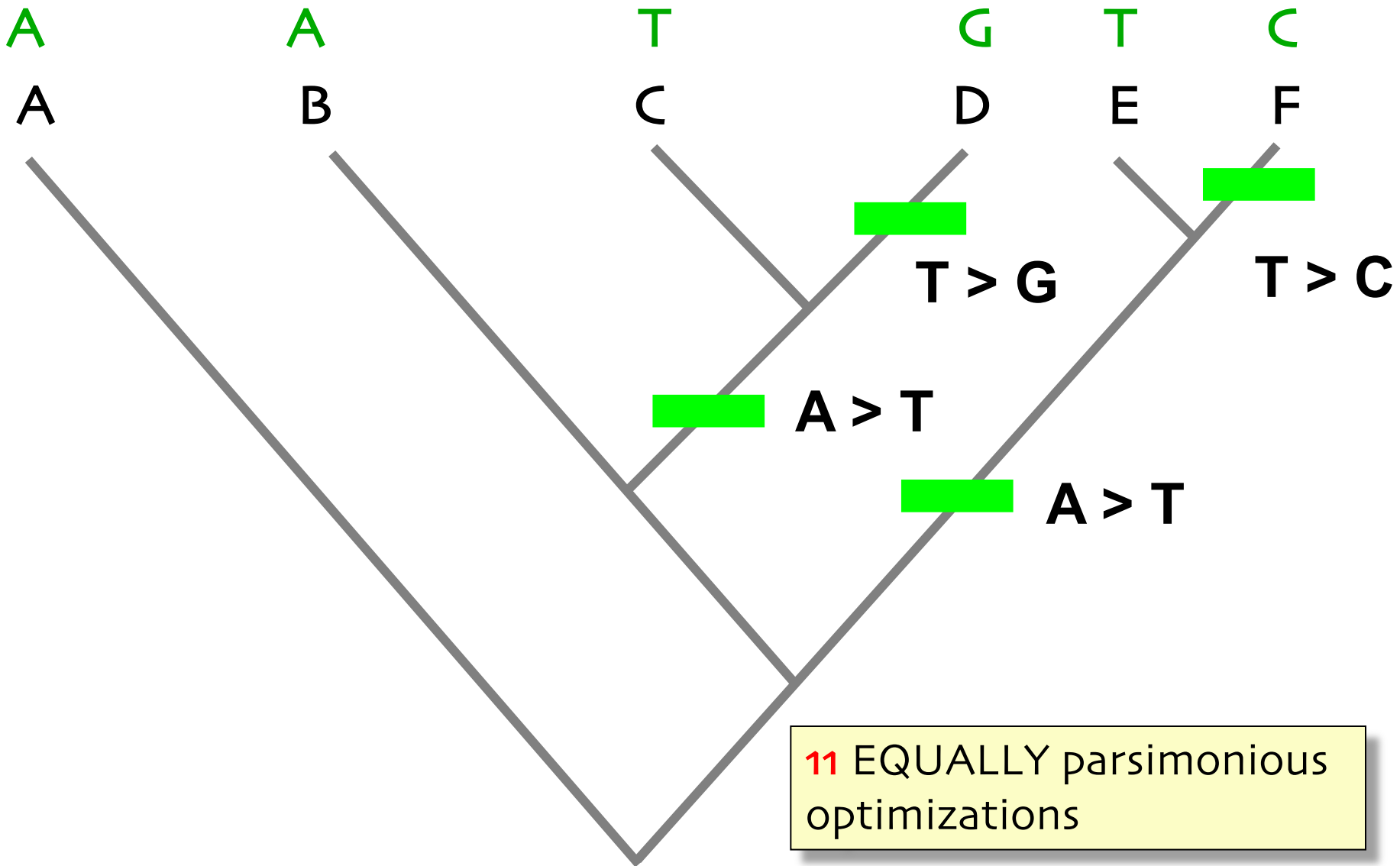


A = 0  
C = 1  
G = 2  
T = 3











# Wagner optimization

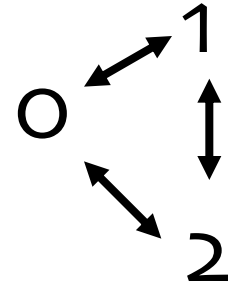
Farris, J.S. 1970. Methods for computing Wagner trees.

*Systematic Zoology* 19: 83-92.

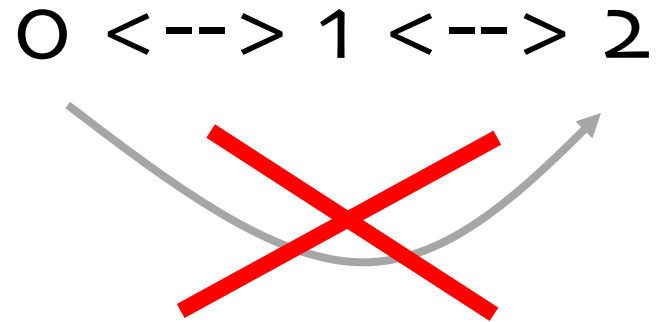
Fitch, W.M. 1971. Toward defining the course of evolution : minimal change for a specific tree topology.

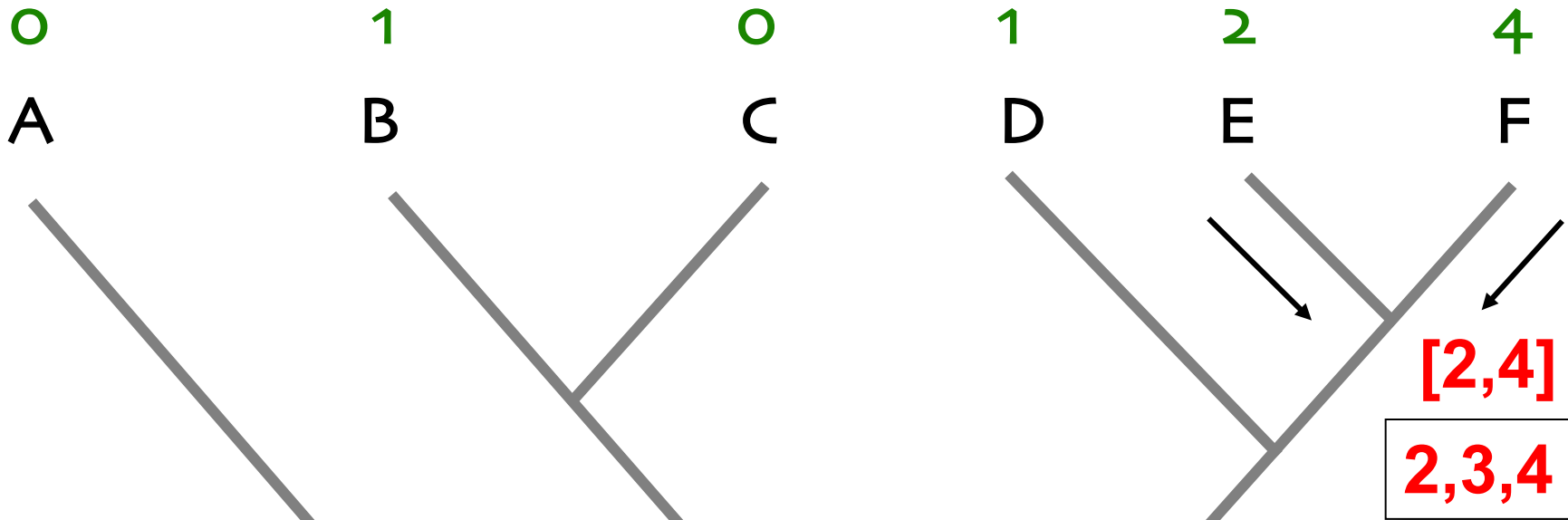
*Systematic Zoology* 20: 406-416.

FITCH PARSIMONY



WAGNER PARSIMONY



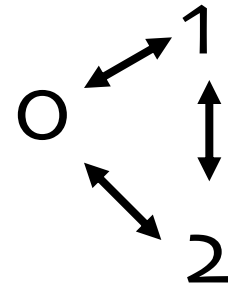


RULE 1: if terminals share character state this will be marked also for their ancestor (intersection,  $\cap$ )

RULE 2: if terminals do not share ch. states (intersection,  $\cap = \emptyset$ ) assign smallest closed interval between states of terminals for their ancestor, i.e.  
 $[a, b] = \{x | a \leq x \leq b\}$

**DOWNWARD  
PASS**

# FITCH PARSIMONY



$$A = \{0,1,2,3\}$$

$$A \cap B = \emptyset$$

$$B = \{5,6\}$$

$$A \cup B = \{0,1,2,3,5,6\}$$

# WAGNER PARSIMONY



$$A = \{0,1,2,3\}$$

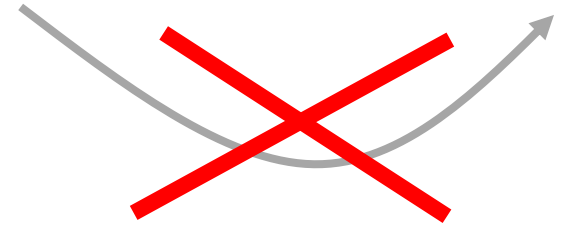
$$A \cap B = \emptyset$$

$$B = \{5,6\}$$

$$A \cup B = [3,5] = \{x \mid 3 \leq x \leq 5\}$$

# WAGNER PARSIMONY

0 <--> 1 <--> 2

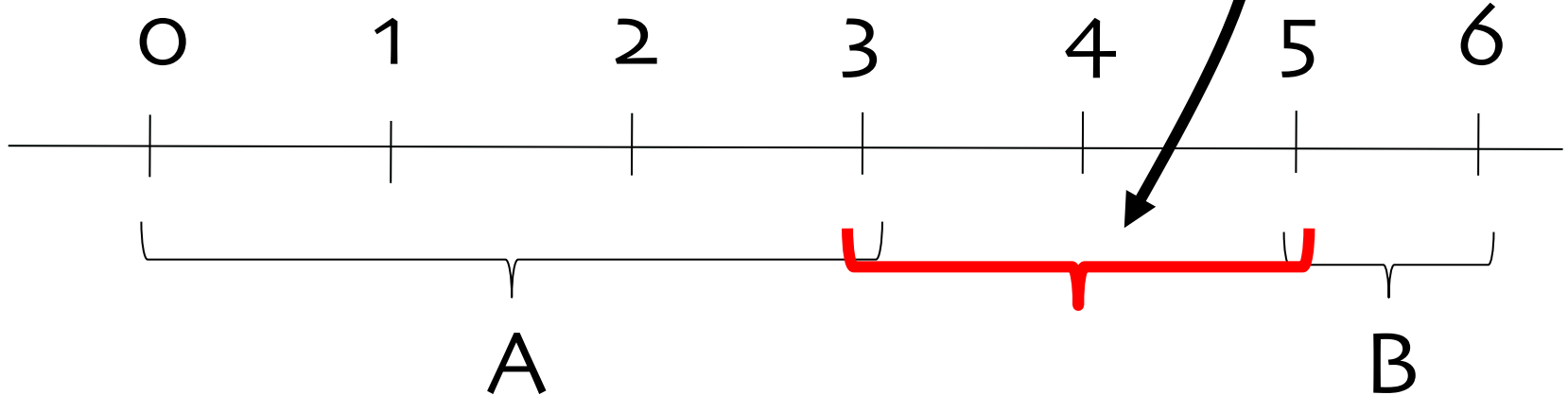


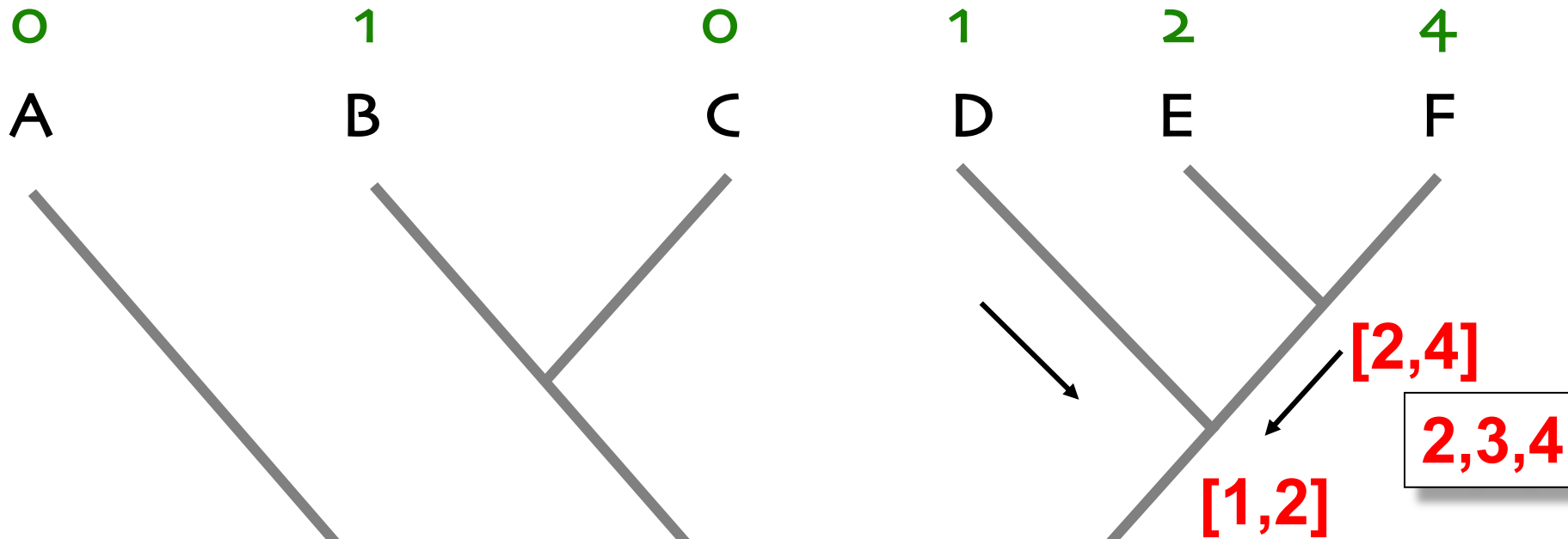
$$A = \{0, 1, 2, 3\} \quad A \cap B = \emptyset$$

$$B = \{5, 6\} \quad A \cup B = \{3, 4, 5\} = \{x \mid 3 \leq x \leq 5\}$$



SMALLEST closed interval between A & B

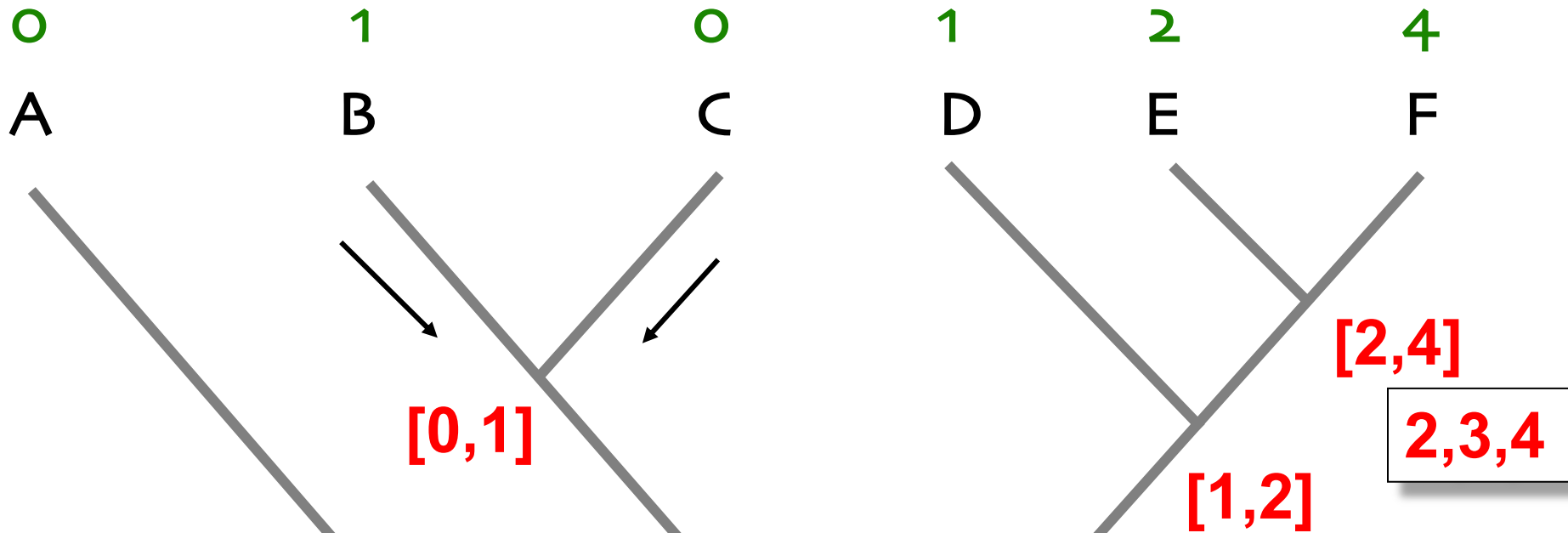




RULE 1: if terminals share character state this will be marked also for their ancestor (intersection,  $\cap$ )

RULE 2: if terminals do not share ch. states (intersection,  $\cap = \emptyset$ ) assign smallest closed interval between states of terminals for their ancestor, i.e.  
 $[a, b] = \{x | a \leq x \leq b\}$

**DOWNWARD  
PASS**

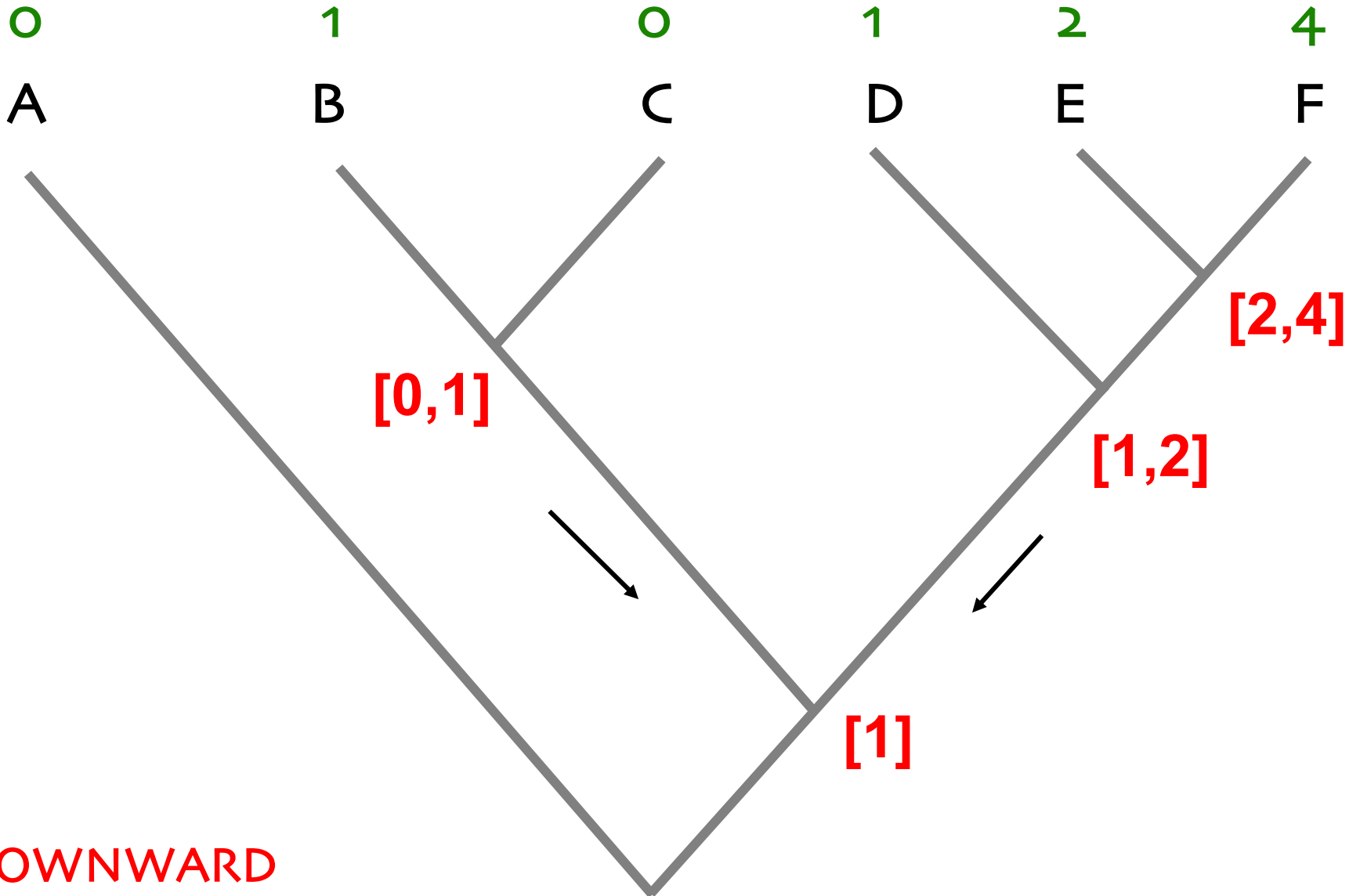


RULE 1: if terminals share character state this will be marked also for their ancestor (intersection,  $\cap$ )

RULE 2: if terminals do not share ch. states (intersection,  $\cap = \emptyset$ ) assign smallest closed interval between states of terminals for their ancestor, i.e.  
 $[a, b] = \{x | a \leq x \leq b\}$

**DOWNWARD  
PASS**

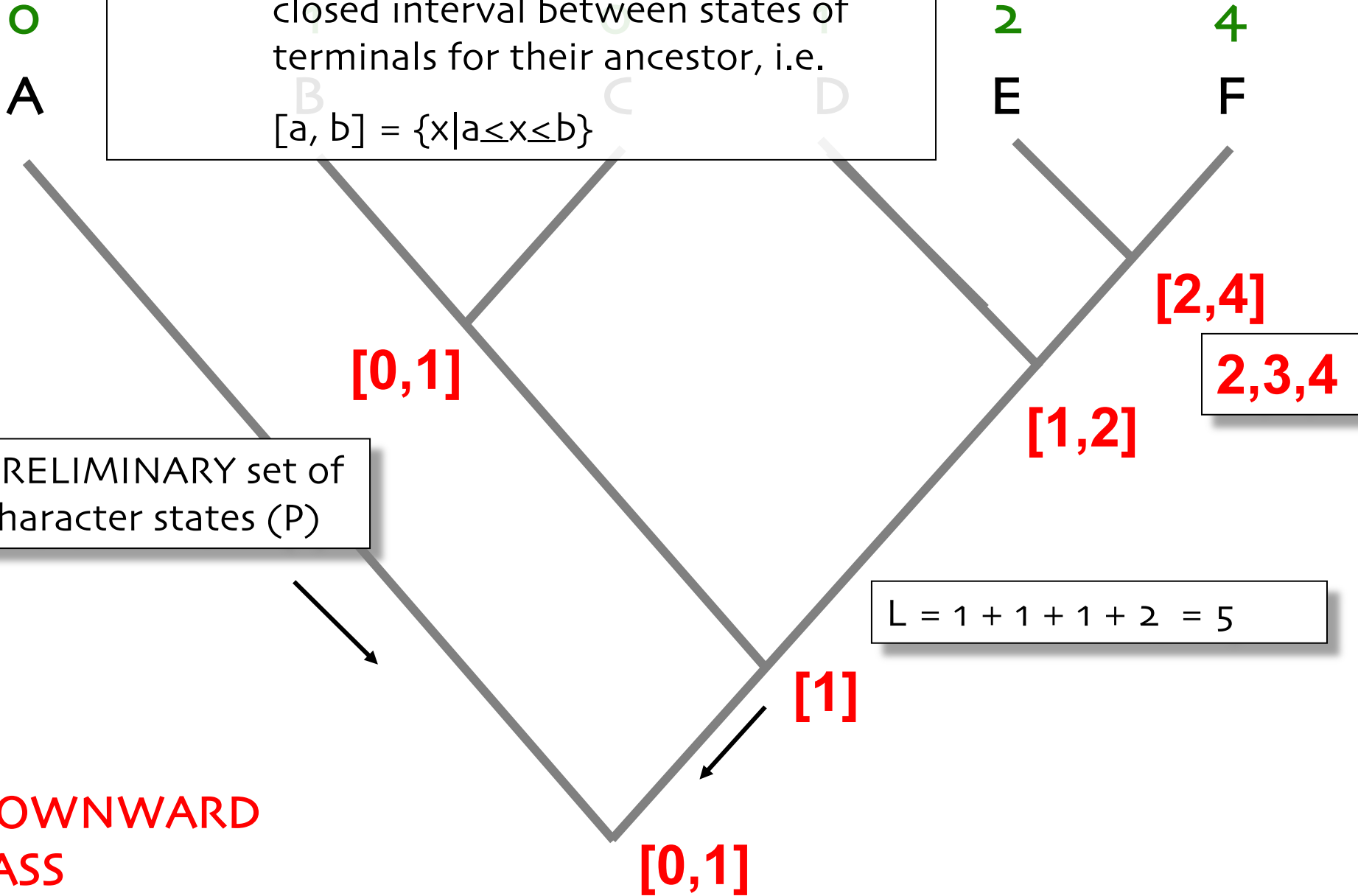
RULE 1: if terminals share character state this will be marked also for their ancestor (intersection,  $\cap$ )

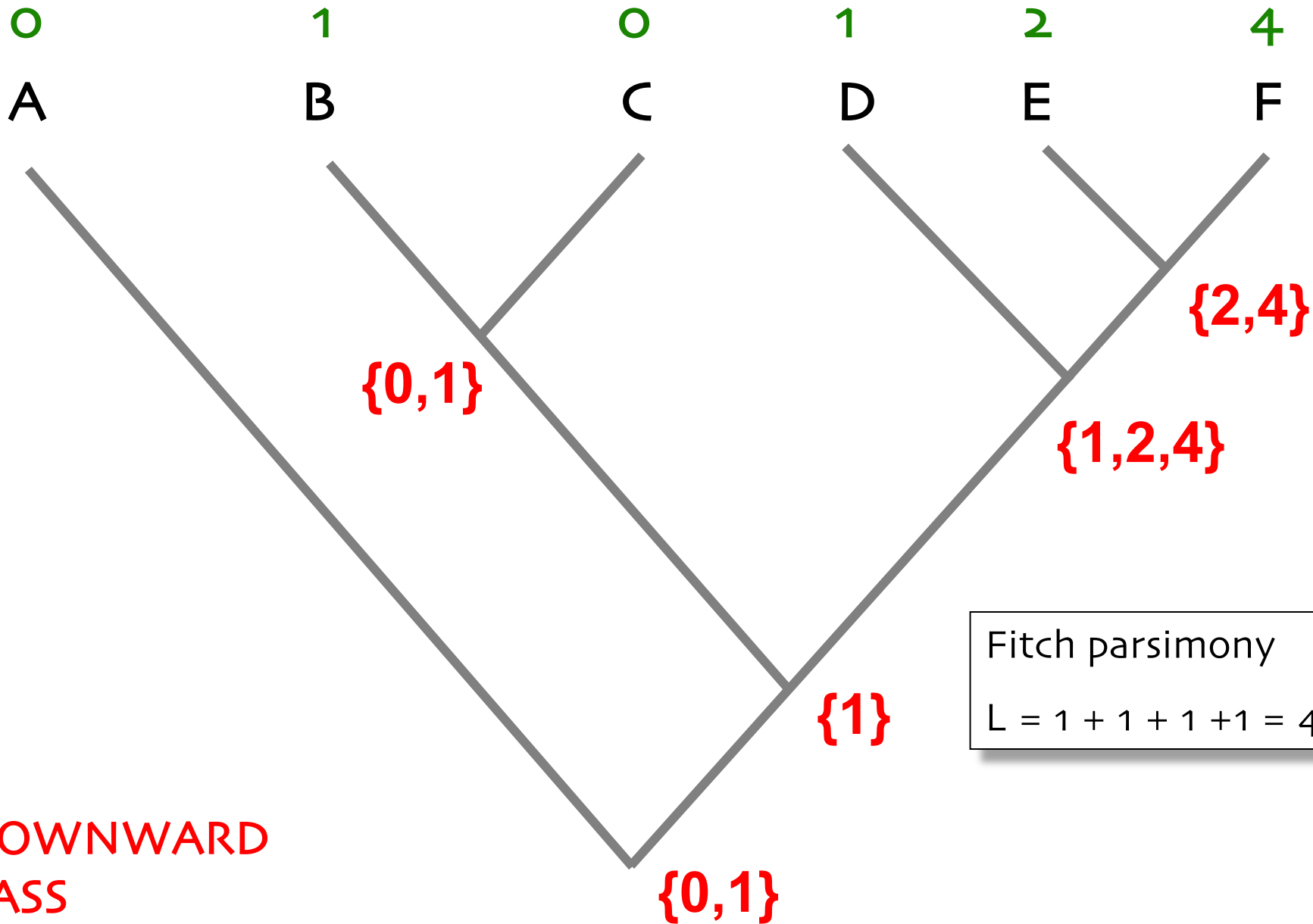


DOWNWARD  
PASS



RULE 2: if terminals do not share ch. states (intersection,  $\cap = \emptyset$ ) assign smallest closed interval between states of terminals for their ancestor, i.e.  
 $[a, b] = \{x | a \leq x \leq b\}$





“ upward pass” rules (Goloboff 1993)

- *PRELIMINARY*( $P$ ) state set for root and terminals is their final set

$$(P = F)$$

$A$ , character state of  
immediate ancestor

RULE 1. If  $A \cap P = A$ ,  $F = A$ .

RULE 2. If rule 1 does not apply, and  $(L \cup R) \cap A \neq \emptyset$ , define  $X$  as  
 $X = (L \cup R \cup P) \cap A$ . If  $X \cap P \neq \emptyset$ ,  $F = X$ . If  $X \cap P = \emptyset$ ,  $F$  equals  
the **LARGEST** closed interval between  $X$  and state in  $P$  closest to  $X$ .

RULE 3. If rules 1 & 2 do not apply,  $F$  equals the **LARGEST** closed interval  
between the state in  $P$  closest to  $A$  and the state in  $(L \cup R)$  closest  
to  $A$ .

---

**Algorithm 9.2: AdditiveUpPass**

---

**Data:** Tree,  $T = (V, E)$ , with leaf taxa  $L \subset V$

**Data:** Character set,  $c \in \mathbb{N}$  for  $V$  initialized by algorithm [9.1](#).

**Data:** Preliminary states are  $c^p$ , *final* or up-pass states are  $c^f$ .

**Data:** Vertex,  $v$ , of  $T$ , initially the root. If  $v \notin L$  then  $v$  has two descendants  $v_{left}$  and  $v_{right}$ .

**Result:** Require that  $c$  contain the set of all states consistent with the tree cost

leaf or root;

if  $v \in L$  or  $v = root$  then

    |  $c_v^f \leftarrow c_v^p$ ;

if  $v \notin L$  then

    | if  $c_v^p \cap c_{v_{parent}}^f = c_{v_{parent}}^f$  then

        |  $c_v^f \leftarrow c_{v_{parent}}^f$ ;

    | else if  $(c_{v_{left}}^p \cup c_{v_{right}}^p) \cap c_{v_{parent}}^f \neq \emptyset$  then

        |  $X = (c_{v_{left}}^p \cup c_{v_{right}}^p \cup c_v^p) \cap c_{v_{parent}}^f$ ;

        | if  $X \cap c_v^p \neq \emptyset$  then

            |  $c_v^f \leftarrow X$ ;

        | else

            | Largest closed interval between  $X$  and  $c_v^p$

            |  $c_v^f \leftarrow lci\{X, c_v^p\}$ ;

    | else

        | Largest closed interval between  $\{c_v^p$  and  $c_{v_{parent}}^f\}$

        | and  $\{(c_{v_{left}}^p \cup c_{v_{right}}^p)$  and  $c_{v_{parent}}^f\}$

        |  $c_v^f \leftarrow$

        |  $lci\{[c_v^p \text{ closest to } c_{v_{parent}}^f], [(c_{v_{left}}^p \cup c_{v_{right}}^p) \text{ closest to } c_{v_{parent}}^f]\}$

    | Recurse up the tree until all  $V \notin L$  are updated

    |  $AdditiveUp(T, c, v_{left})$ ;

    |  $AdditiveUp(T, c, v_{right})$ ;

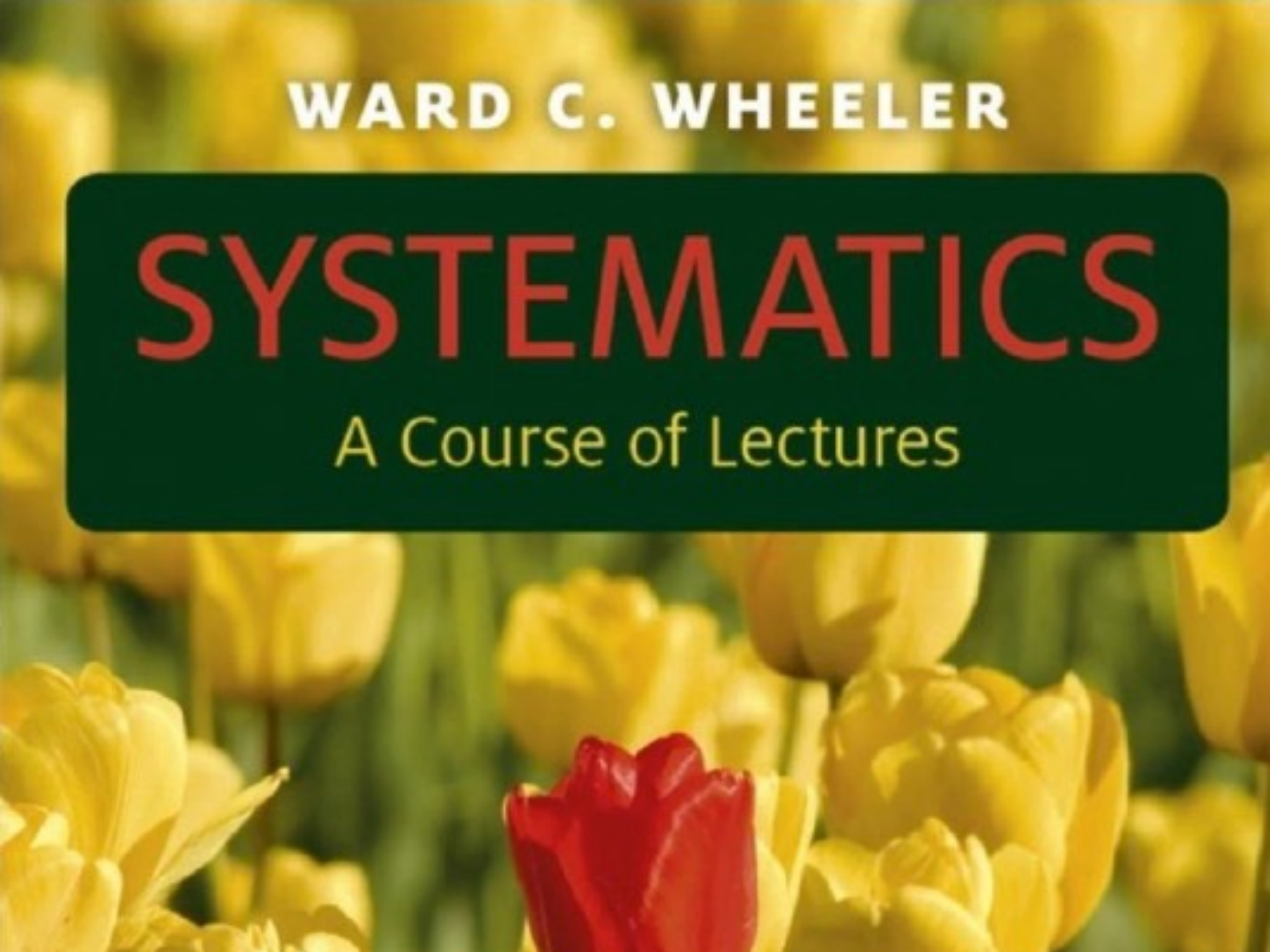
---

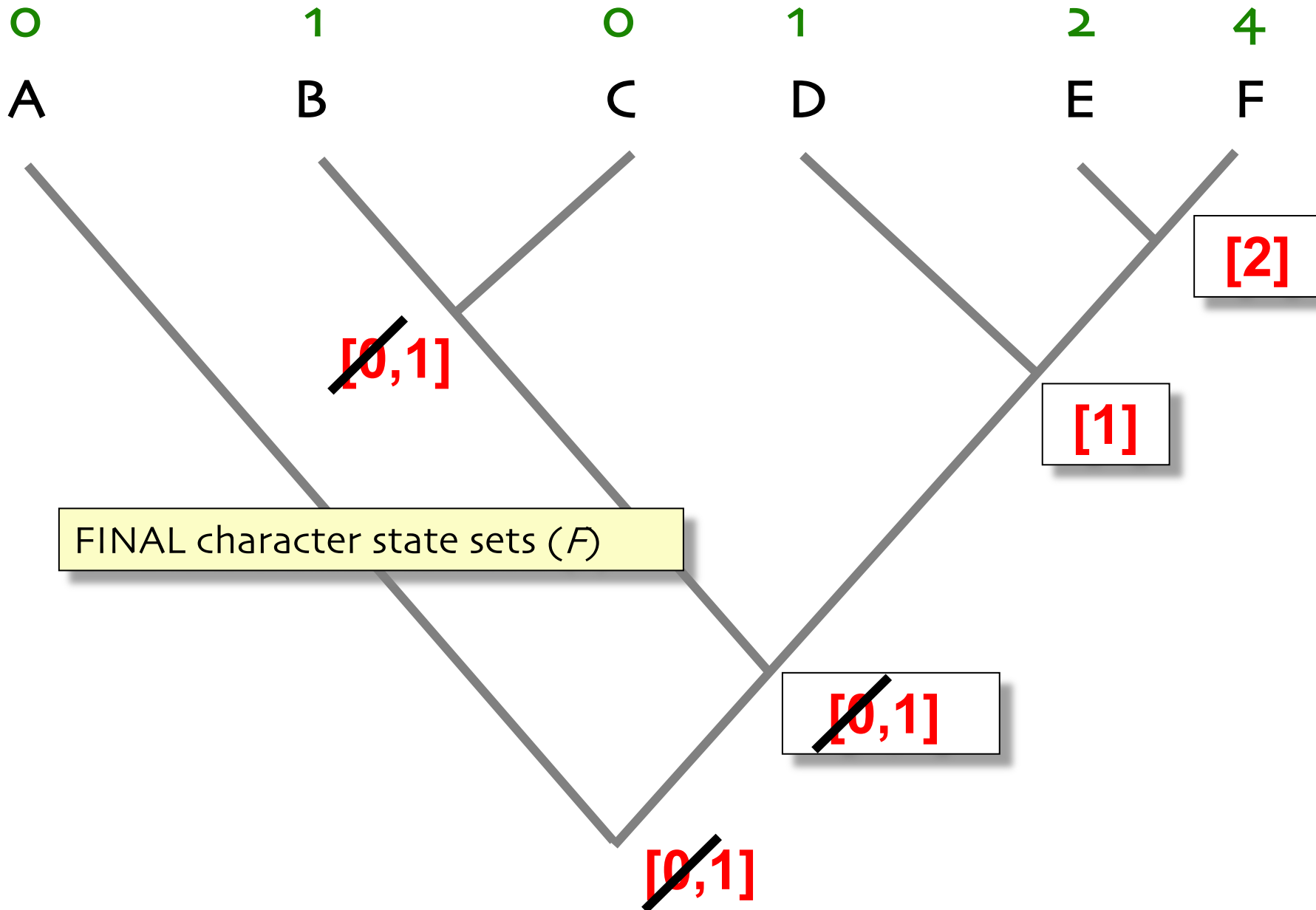
Wheeler 2012

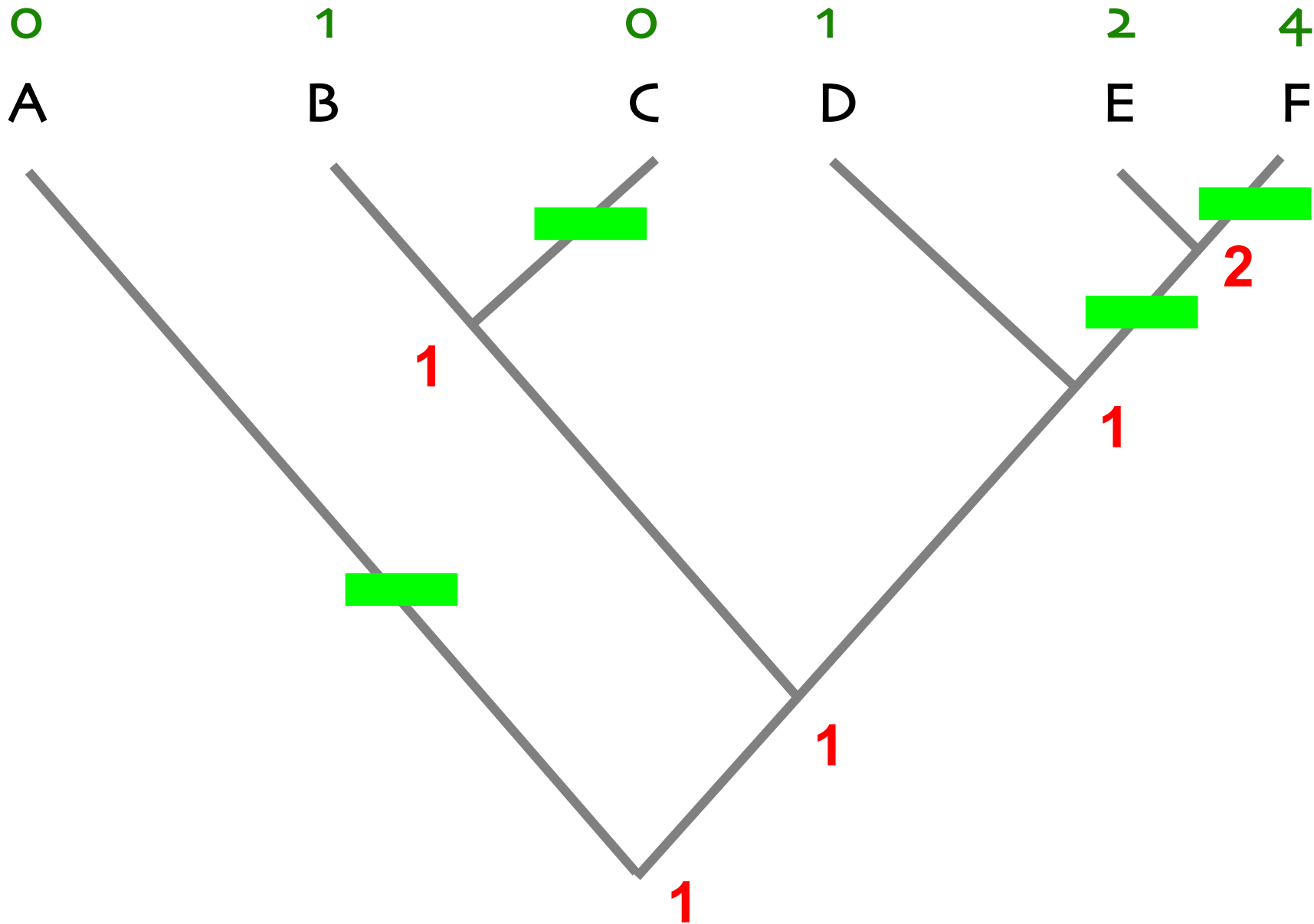
WARD C. WHEELER

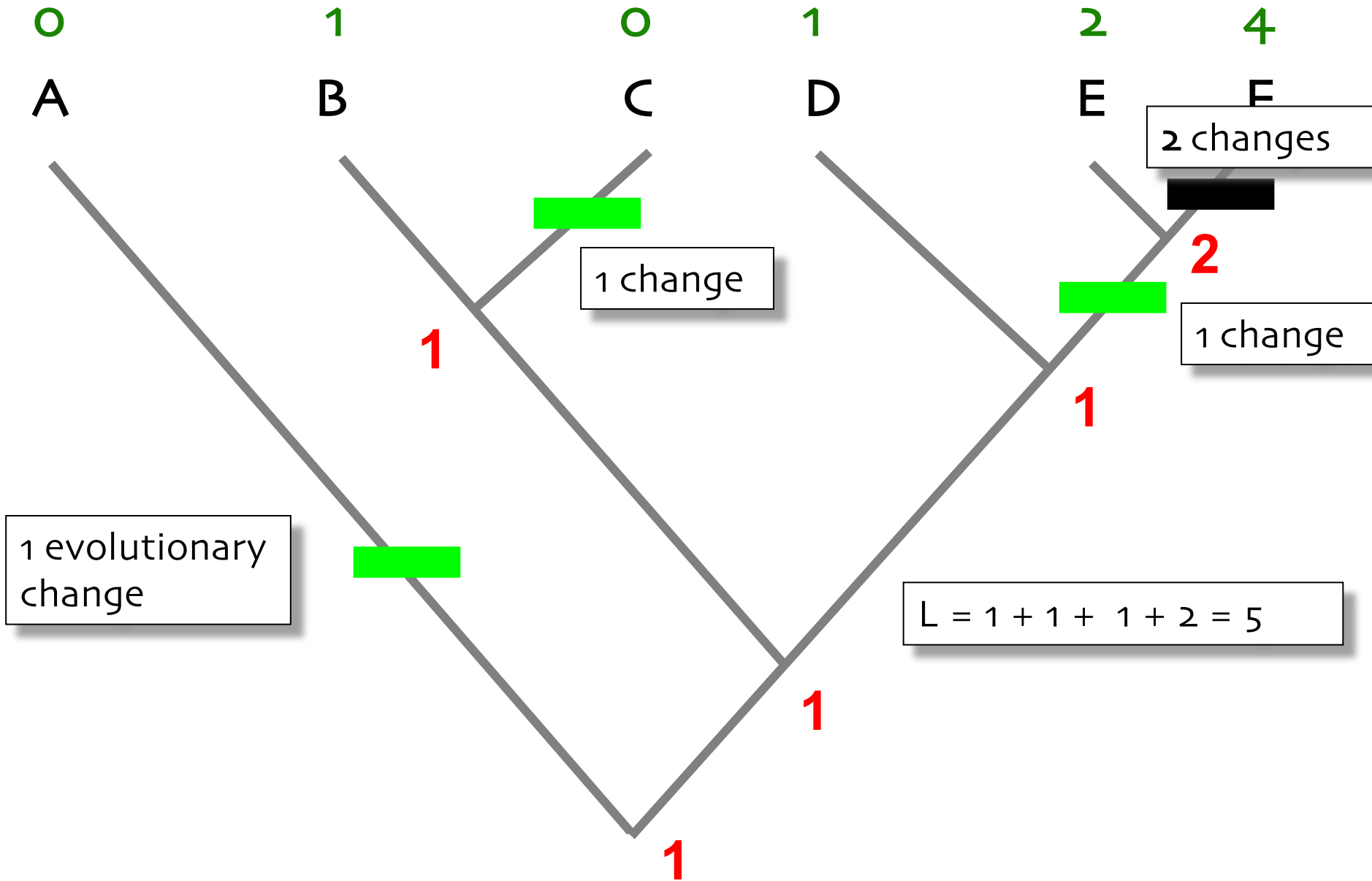
# SYSTEMATICS

A Course of Lectures

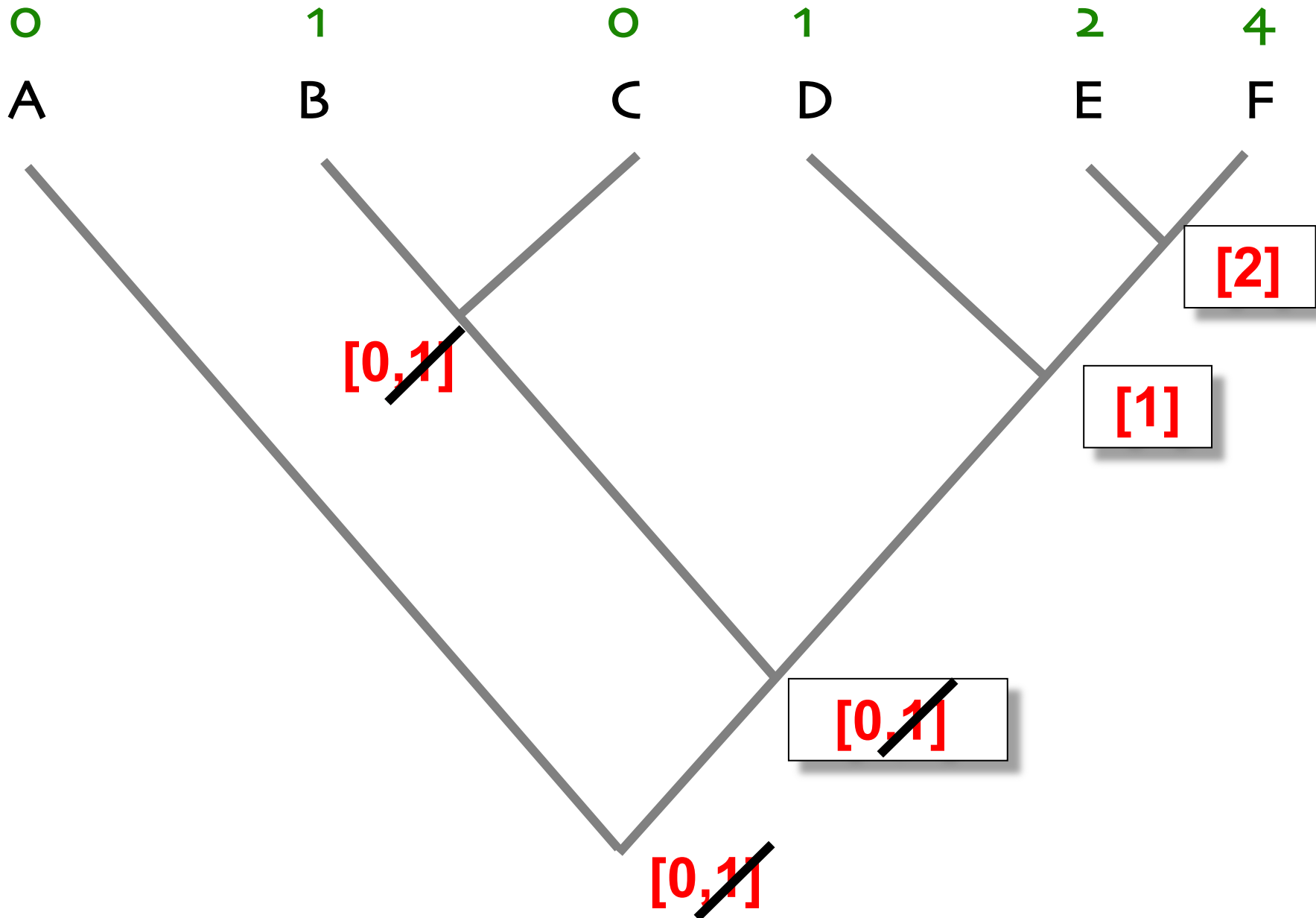


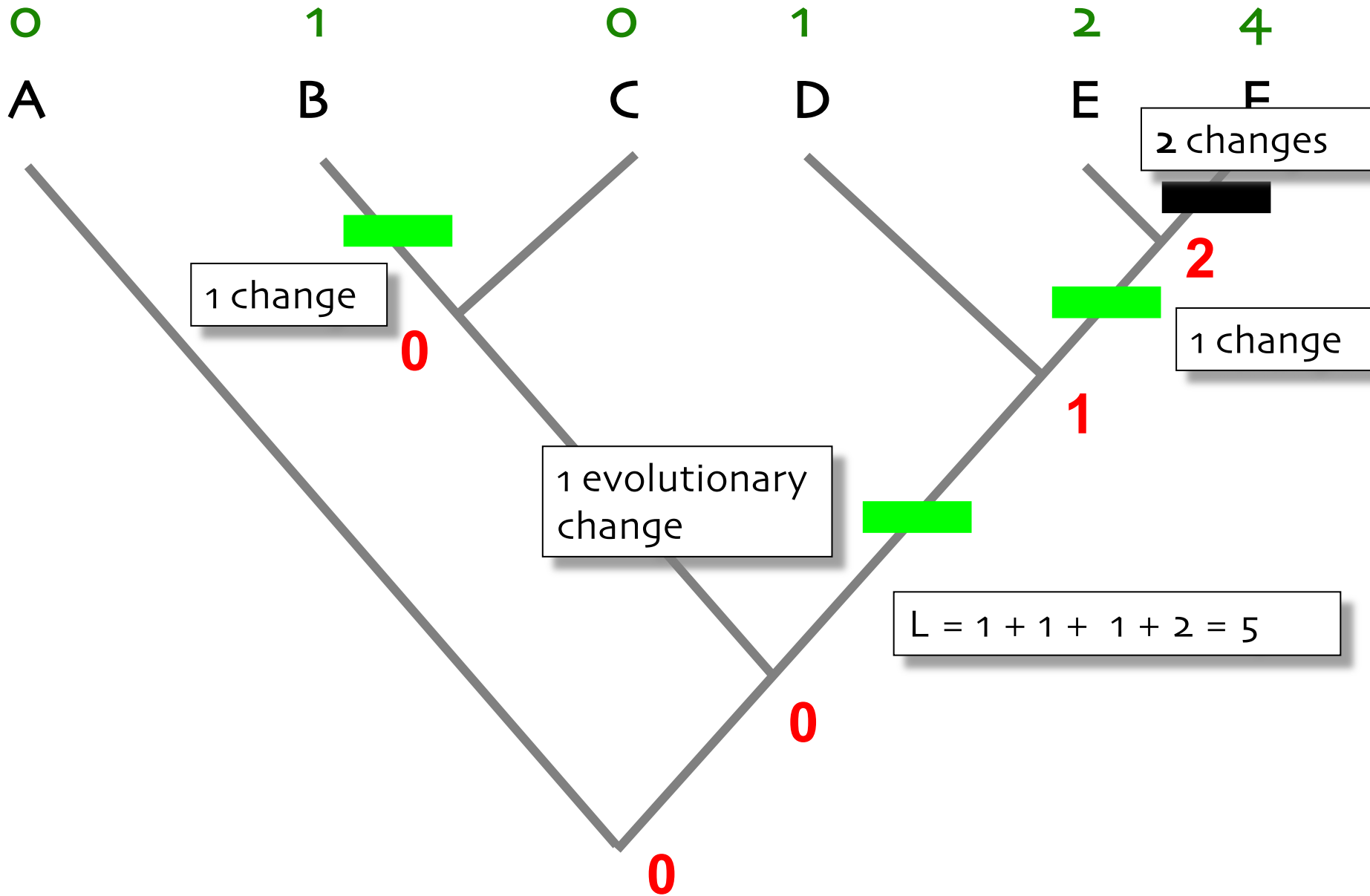












# OPTIMIZATION

Goloboff, P. 1993. Character optimization and calculation of tree lengths. *Cladistics* 9: 433-436.

programs to find ALL equally parsimonious character state reconstructions

MacClade, Mesquite

Swofford, D. L. & Maddison, W. P. 1987. Reconstructing ancestral character states under Wagner parsimony. *Mathematical Biosciences* 87: 199-229.

...frequently ... we are interested not only in the branching pattern but also in the evolutionary hypothesis: a phylogeny coupled with the reconstructed states of the characters in the hypothetical ancestors...

...when multiple, equally parsimonious character-state reconstructions exist, we must be careful in interpreting **any ONE** solution...

# OPTIMIZATION

HTU, Hypothetical Taxonomic Unit

possible character state hypotheses for internal nodes (HTU)

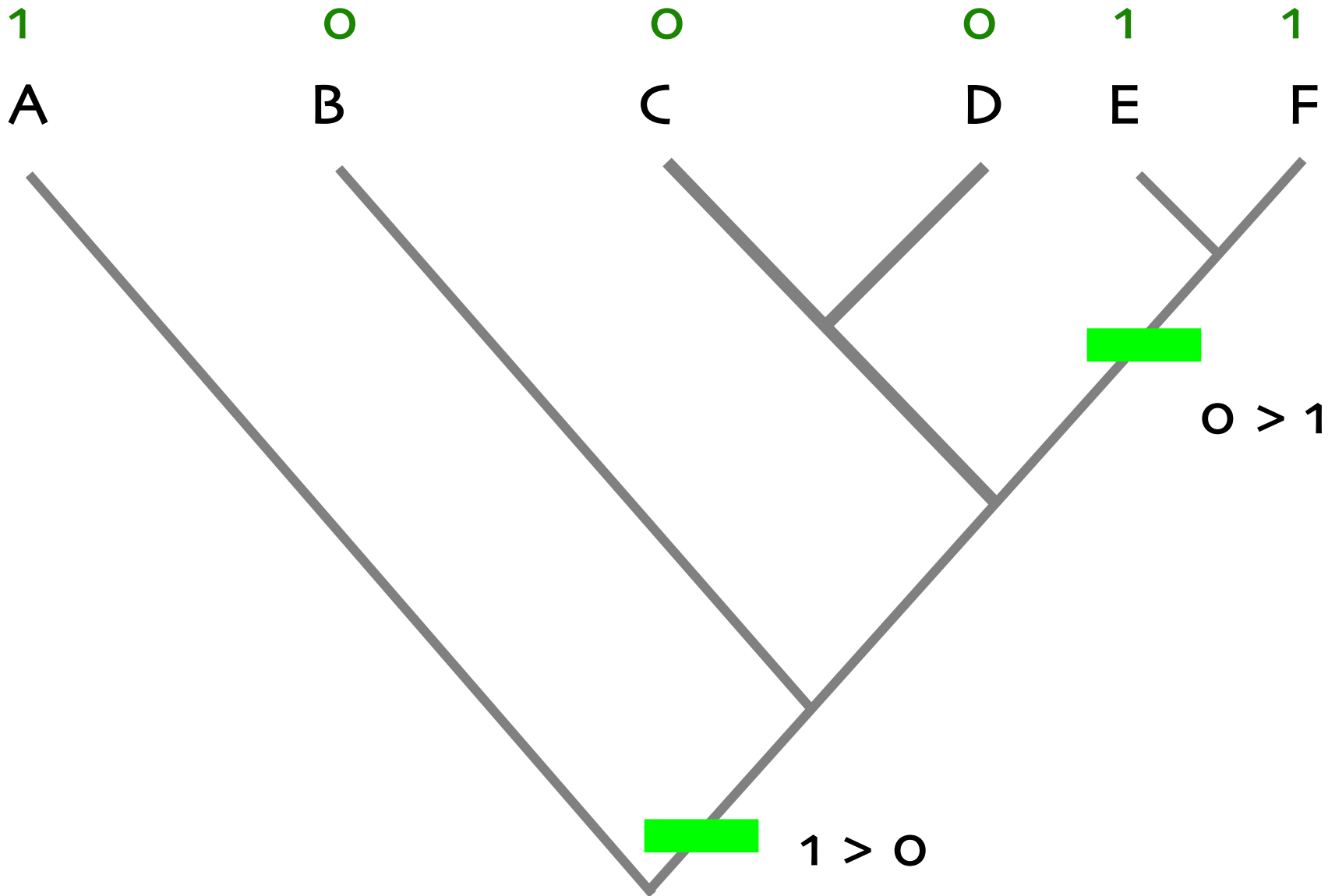
MPR, Most Parsimonious Reconstruction set

practical consequences of different reconstructions?

# ACCTRAN optimization

ACCellarated TRANSformation

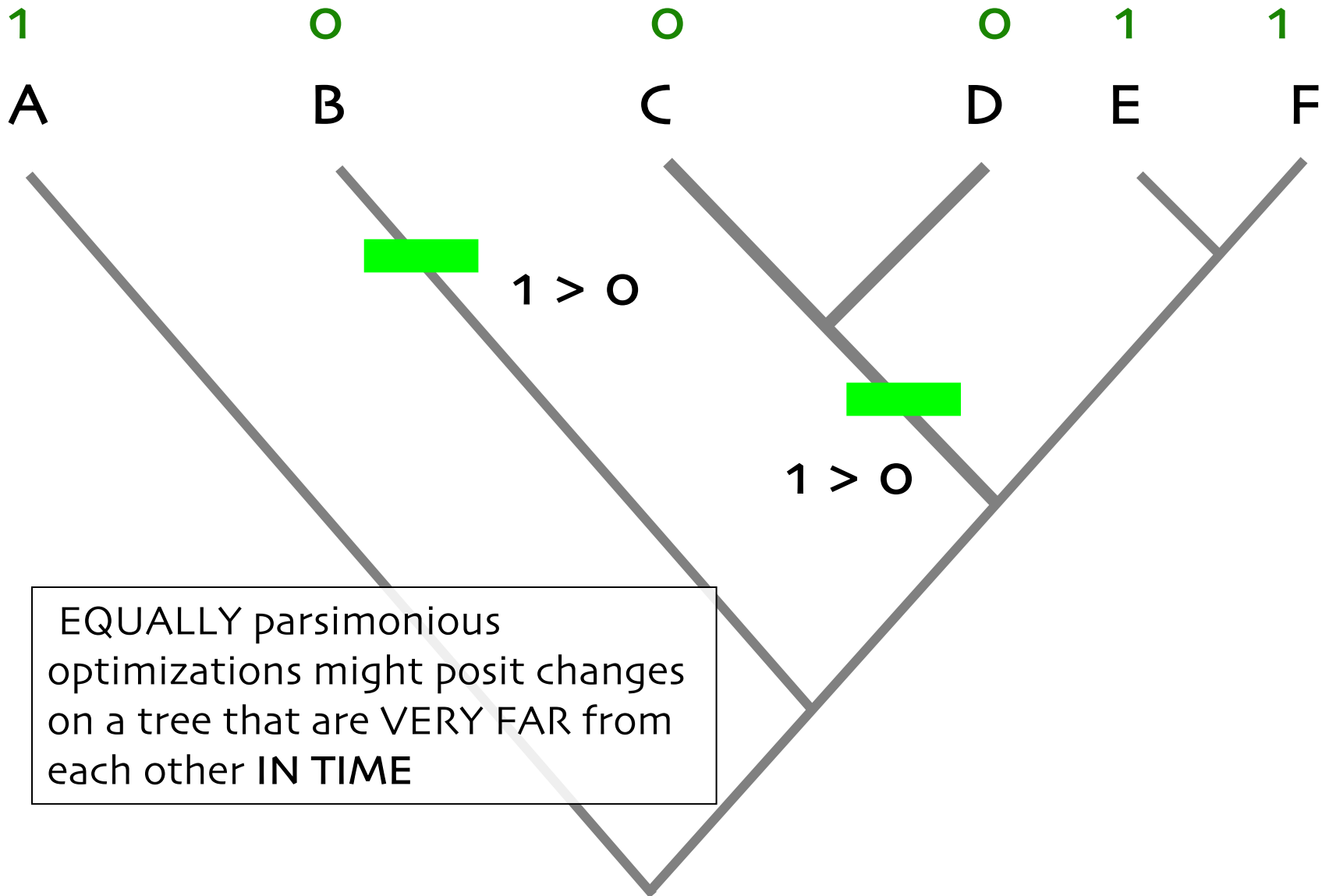
favors reversals, changes are assumed to have taken place as early as possible



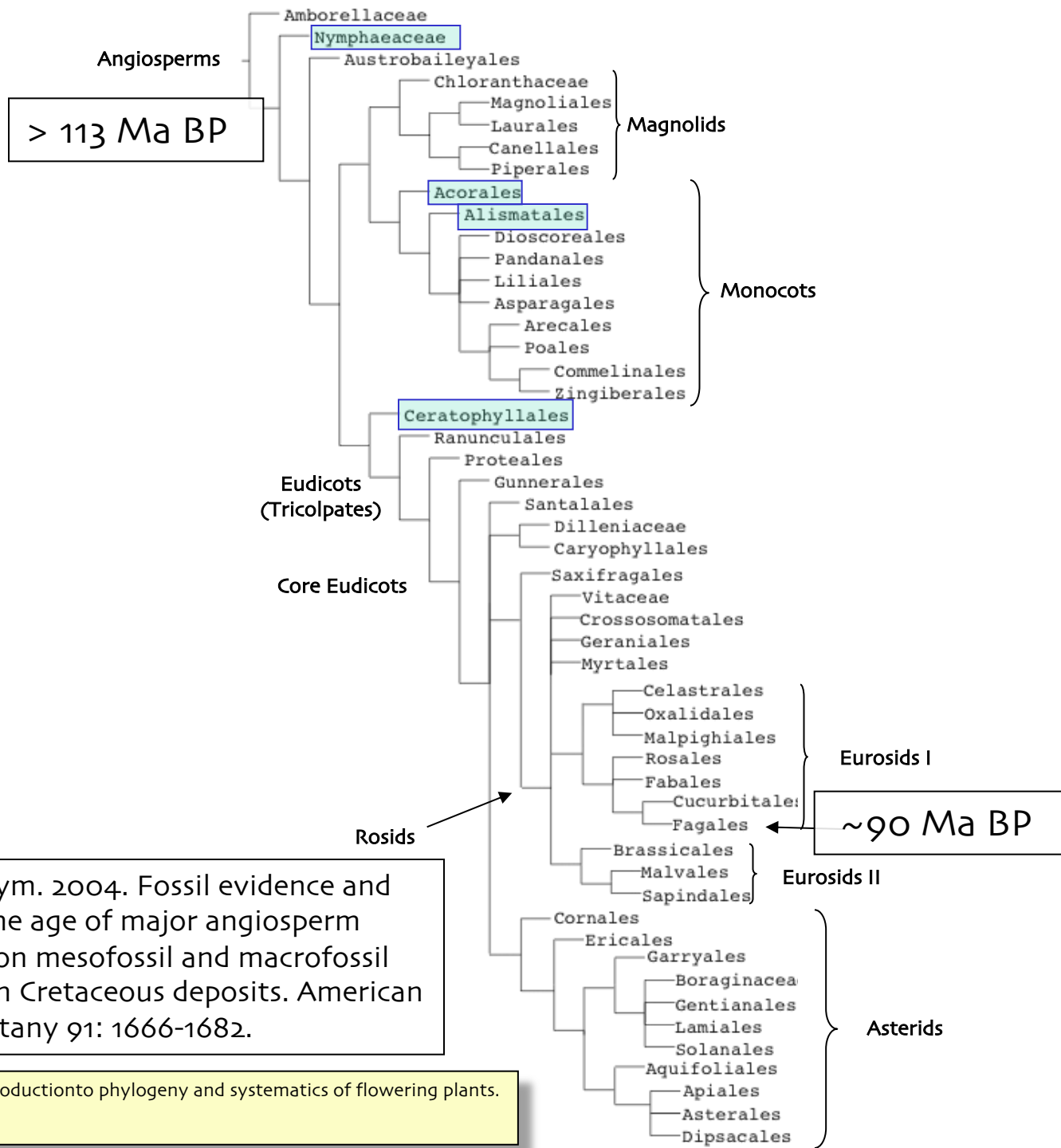
# DELTRAN optimization

DELayed TRANSformation

favors parallelism, changes are assumed to have taken place as late as possible







Crepet, W.L. ym. 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.

Bremer, K. ym. 2003. Introduction to phylogeny and systematics of flowering plants. Acta Univ. Upsal. 33: 2.

# Wagner optimization

Farris, J.S. 1970. Methods for computing Wagner trees.  
*Systematic Zoology* 19: 83-92.

Fitch parsimony	
Wagner	"
Dollo	"
Camin-Sokal	"
Sankoff	"

Fitch, W.M. 1971. Toward defining the course of evolution :  
minimal change for a specific tree topology.  
*Systematic Zoology* 20: 406-416.

# 3.xi.



1. character optimization

- 2. introduction to direct optimization**

3. summary

# DNA-sequences

- empirical **observations** of the nucleotide order  
of the sequence under study

## alignment

- 1) sequences under comparison are aligned based on guide-tree (hypotheses about **homology at the level of nucleotides**)
- 2) phylogenetic analysis (kind of test of individual homology hypotheses)

## direct optimization

alignment & phylogenetic analyses  
performed simultaneously



hypothesis of phylogeny

**ACTTC CGAATTTGGCT**

**ACTCGATTGCCT**

**ACTTCCGAATTGG-CT**

**||| ||| ||| ||**

**ACT--CGA--TTG-CCT**

**ACTTCCGAATTGGCT**

**||| \* ||| | \* ||**

**ACTC---GATT-GCCT**

**ACTTCCGAATTGGCT**

**||| | ||| | \* ||**

**ACT-C-GA-TT-GCCT**

# Problems & challenges in alignment

```
AACGGTTAAGGTACGGAGAATTAGGCAACCCTAGGATGCGATGCGCAGAGTTAGGTACGGAGAATTAGGCAACCCTAGGAT  
AACGGTTAAGGTACGGAGAATTAGGCAACCCTAGGATGCGATGCGCAGAGTTAGGTACGGAGAATTAGGCAACCCTAGGATG
```

alignment is difficult & problematic >

“unalignable” sequences are ignored from analyses

part of the information LOST

Gatesy & al. 1994. *Mol. Phyl. Evol.* 2: 152-157

large investments in GETTING this material in the first place

length differences of sequences DO INCLUDE historical information

# Problems & challenges in alignment

```
AACGGTTAAGGTACGGAGAATTAGGCAACCCTAGGATGCGATGCGCAGAGTTAGGTACGGAGAATTAGGCAACCCTAGGAT
AACGGTTAAGGTACGGAGAATTAGGCAACCCTAGGATGCGATGCGCAGAGTTAGGTACGGAGAATTAGGCAACCCTAGGATGCG
```

differential “weighting” of different types of transformations **indel & substitutions costs defined**

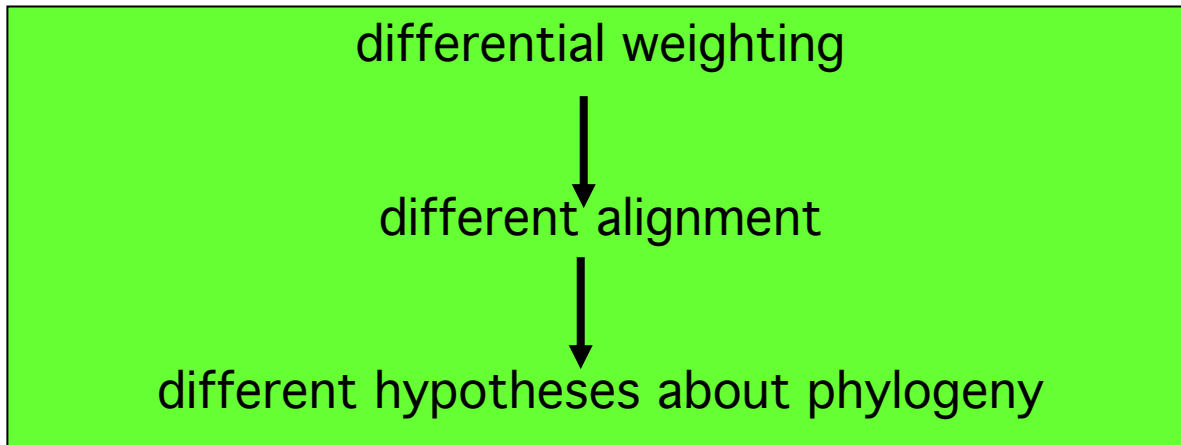
weighting is always *ad hoc*, but it can dramatically affect results



# Problems & challenges in alignment

AACGGTTAAGGTACGGAGAATTAGGCAACCCTAGGATGCGATGCCGAGAGTTAGGTACGGAGAATTAGGCAACCCTAGGAT  
AACGGTTAAGGTACGGAGAATTAGGCAACCCTAGGATGCGATGCCGAGAGTTAGGTACGGAGAATTAGGCAACCCTAGGATG

differential “weighting” of different types of transformations **indel & substitutions costs defined**  
weighting is always *ad hoc*, but it can dramatically affect results



weighting in “manual” alignment is NOT consistent  
weighting in automatic alignment *ad hoc* but at least **explicit**

Gap = 1  
Tv/Ts = 1

Cost = 2

Cost = 3

TTTACTTT  
TTTG-TTT

TTTACTTT  
TTT-GTTT

TTT-ACTTT  
TTTG--TTT

Cost = 4

Cost = 4

Cost = 3

Gap = 2  
Tv = 2  
Ts = 1

Cost = 6

Cost = 6

Gap = 2  
Tv = 1  
Ts = 2

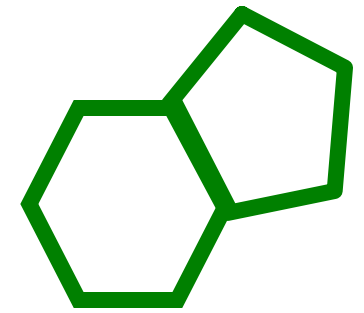
Cost = 3

Gap = 1  
Tv = 2  
Ts = 2

original example by G. Giribet (Harvard Univ.)

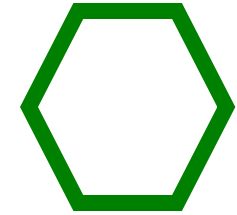
purines: adenine (A)

guanine (G)

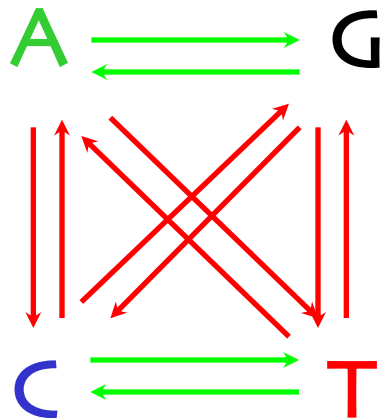


pyrimidins: cytosine (C)

thymine (T)



substitution types:



transitions: 4 types

transversions: 8 types

# DNA-sequences

- empirical **observations** of the nucleotide order  
of the sequence under study

## alignment

- 1) sequences under comparison are aligned  
based on **GUIDE-TREE** (hypotheses  
about **homology at the level of nt's**)
- 2) phylogenetic analysis (kind of test of  
individual homology hypotheses)



hypothesis of phylogeny

# DNA-sequences

- empirical **observations** of the nucleotide order  
of the sequence under study

## alignment

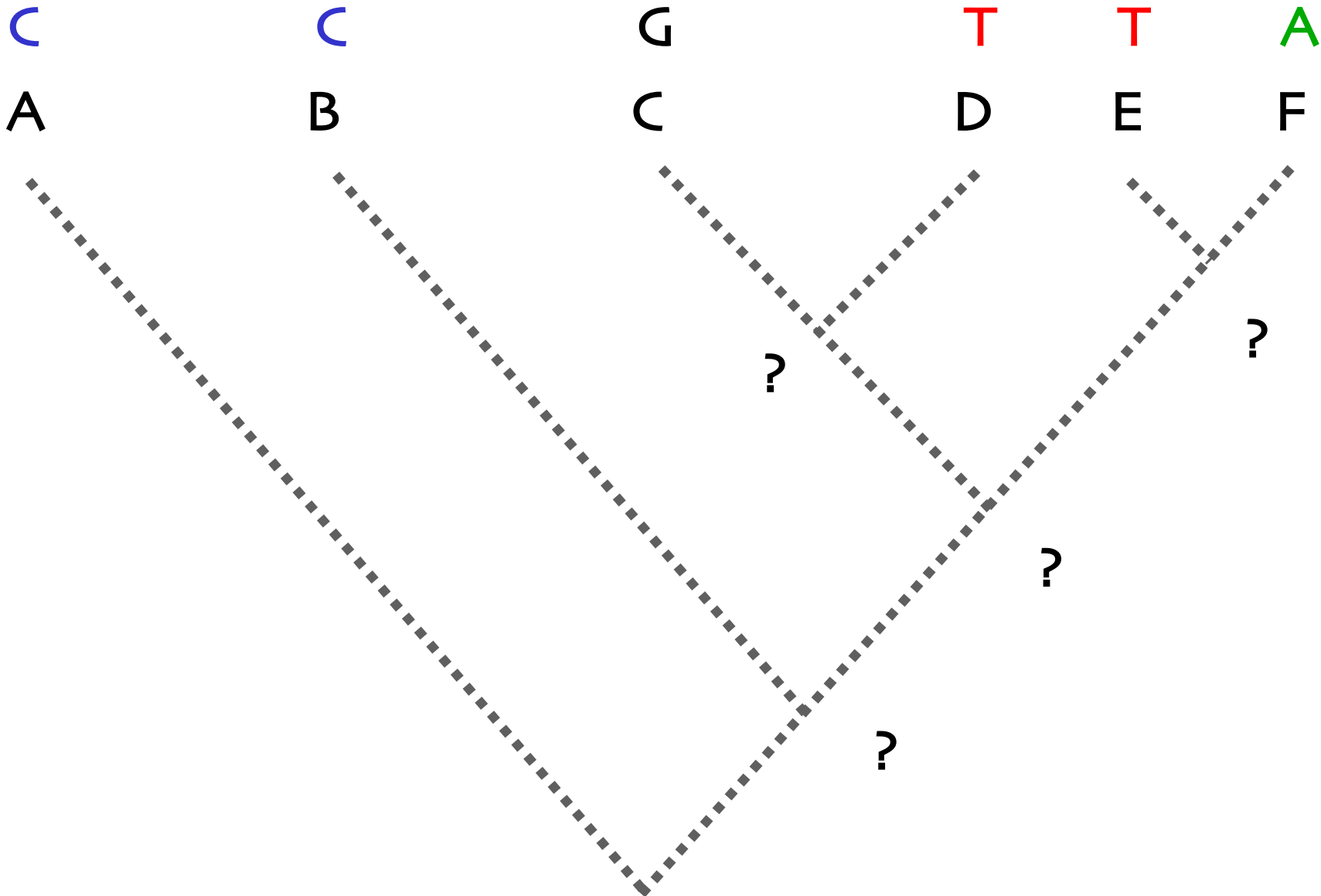
- 1) sequences under comparison are aligned based on guide-tree (hypotheses about homology at the level of nucleotides)
- 2) phylogenetic analysis (kind of test of individual homology hypotheses)

## direct optimization

alignment & phylogenetic analyses  
performed simultaneously

different transformations are treated  
exactly in the same way **ALL THE TIME**

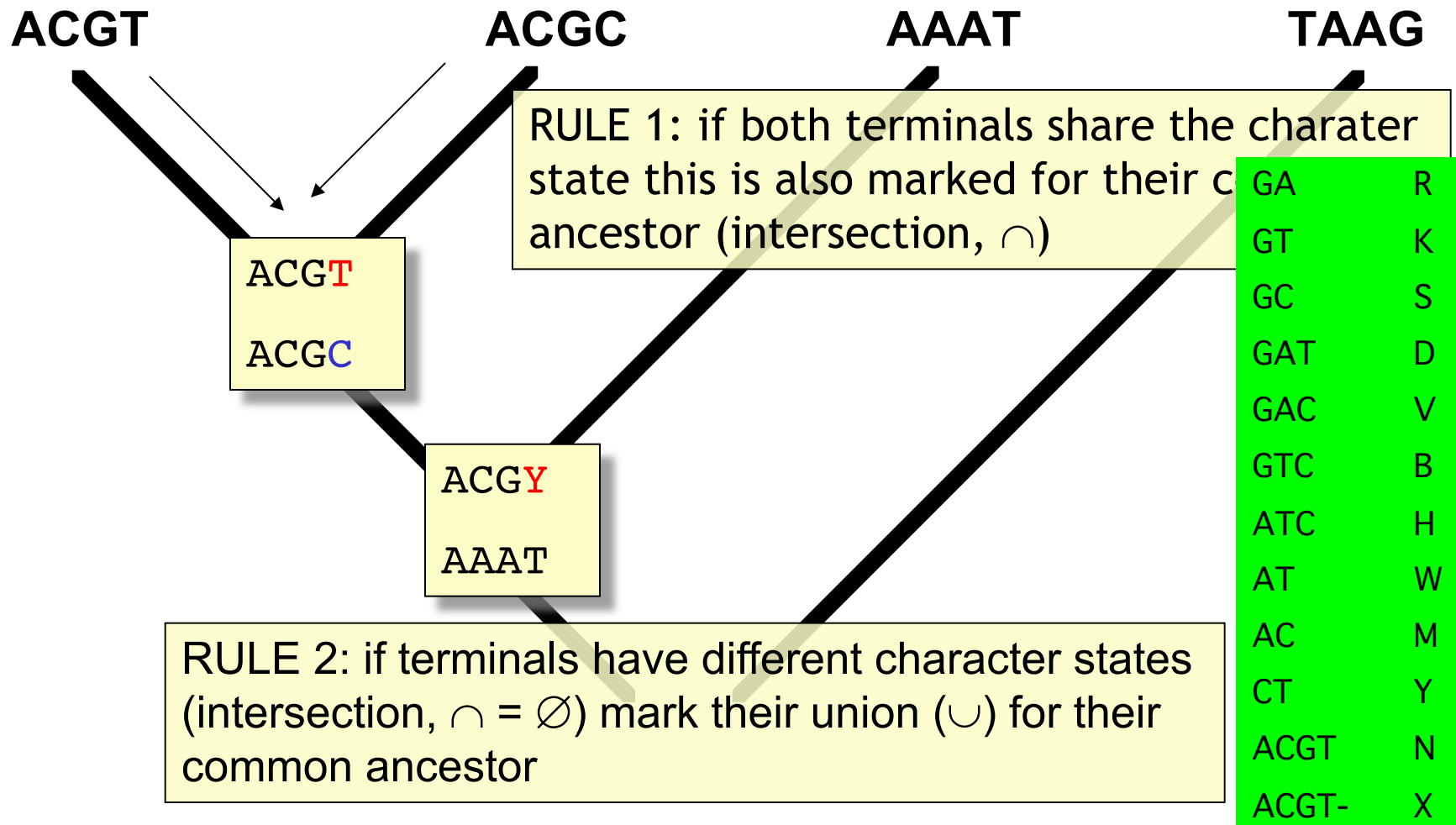
hypothesis of phylogeny



# Direct optimization (DO)

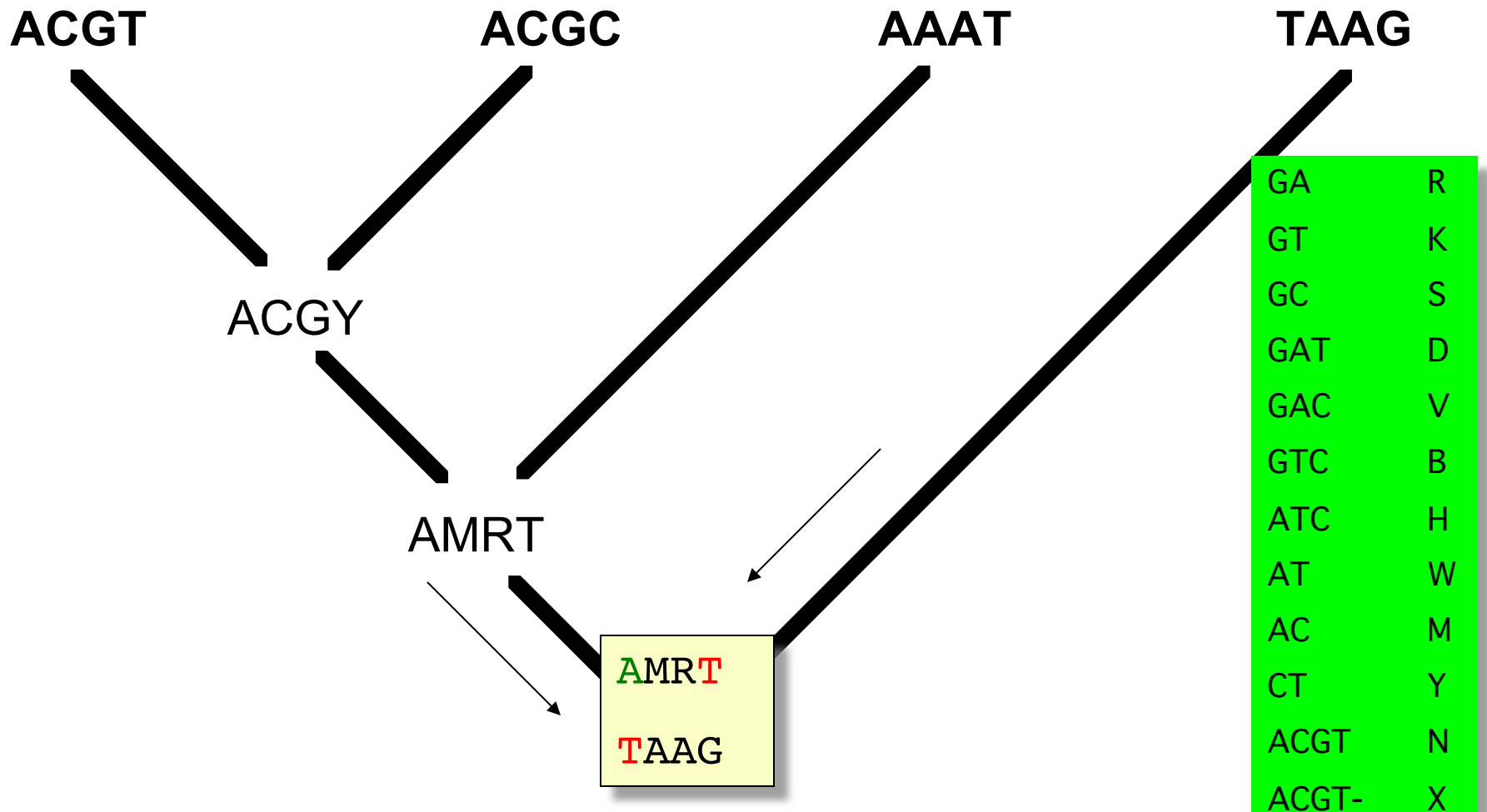
down pass

AACGGTTAAGGTACGGAGAATTAGGCAACCCTAGGATGCGATGCCGAGAGTTAGGTACGGAGAATTAGGCAACCCTAGGAT  
AACGGTTAAGGTACGGAGAATTAGGCAACCCTAGGATGCGATGCCGAGAGTTAGGTACGGAGAATTAGGCAACCCTAGGATGCG



# Direct optimization (DO)

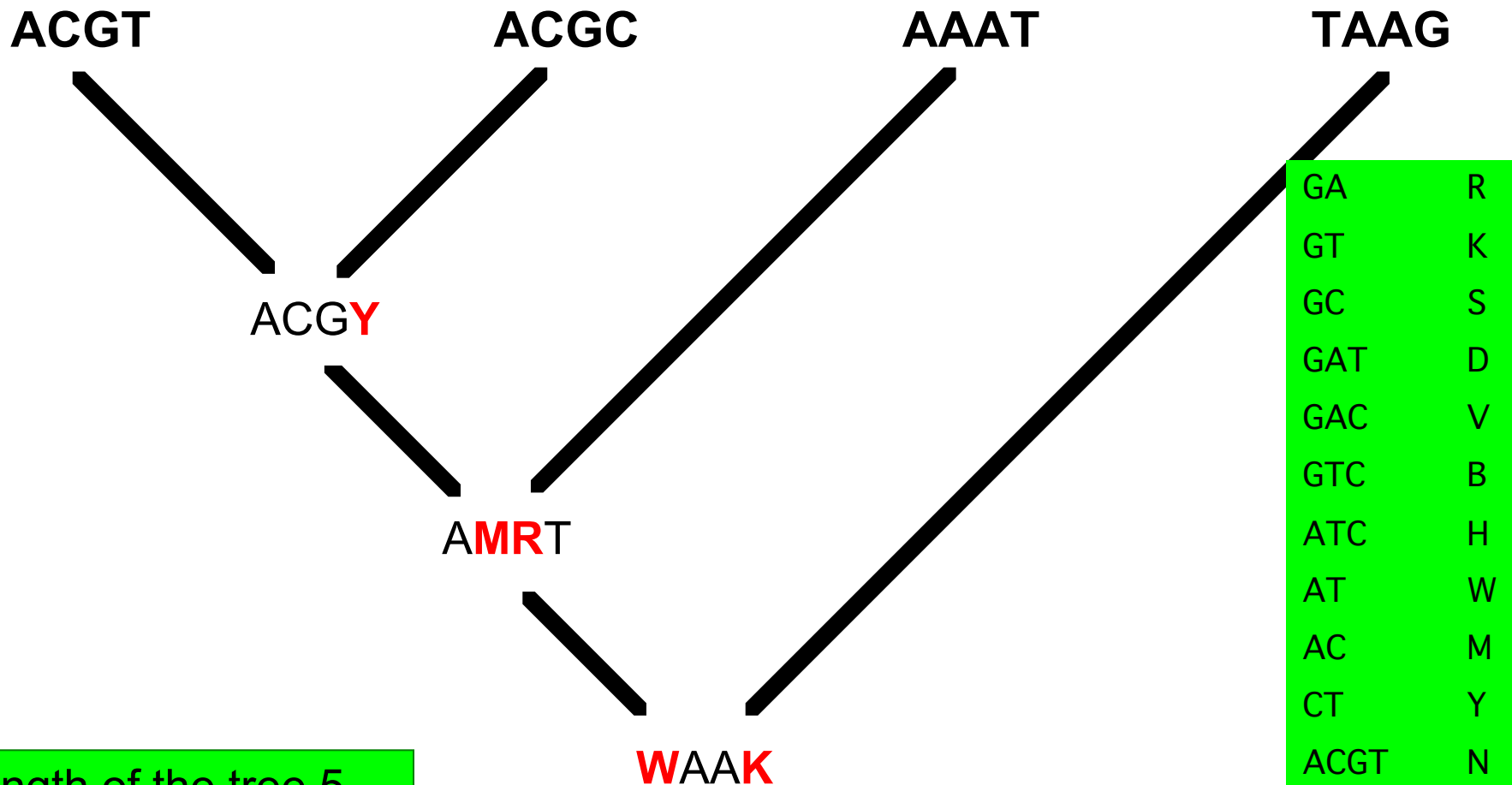
AACGGTTAAGGTACGGAGAATTAGGCAACCCTAGGATGCGATGCGCAGAGTTAGGTACGGAGAATTAGGCAACCCTAGGAT  
AACGGTTAAGGTACGGAGAATTAGGCAACCCTAGGATGCGATGCGCAGAGTTAGGTACGGAGAATTAGGCAACCCTAGGATGCG





# Direct optimization (DO)

AACGGTTAAGGTACGGAGAATTAGGCAACCCCTAGGATGCGATGCGCAGAGTTAGGTACGGAGAATTAGGCAACCCGGTAGGAT  
AACGGTTAAGGTACGGAGAATTAGGCAACCCCTAGGATGCGATGCGCAGAGTTAGGTACGGAGAATTAGGCAACCCCTAGGATGCG



length of the tree 5  
changes

“ upward pass” rules (Fitch 1971, Wheeler 2012)

- root final state set = root preliminary state set ( $F = P$ )

RULE 1. If the overlap of the preliminary state,  $P$ , of the node and its ancestor,  $A$ , is equal to  $A$ , (if  $A \cap P = A$ ) then the final state set,  $F$ , is equal to that of the ancestor ( $F = A$ ).

RULE 2. If Rule 1 does not apply and the union of final/preliminary states of the 2 descendants of the current node (*Left* and *Right*) are equal to preliminary states of the current node ( $P = L \cup R$ ), then  $F = P \cup A$ .

RULE 3. If Rule 1 and 2 do not apply the final state set is the preliminary state set, supplemented by state set that is common to the ancestor and descendants ( $F = P \cup (L \cap A) \cup (R \cap A)$ ).

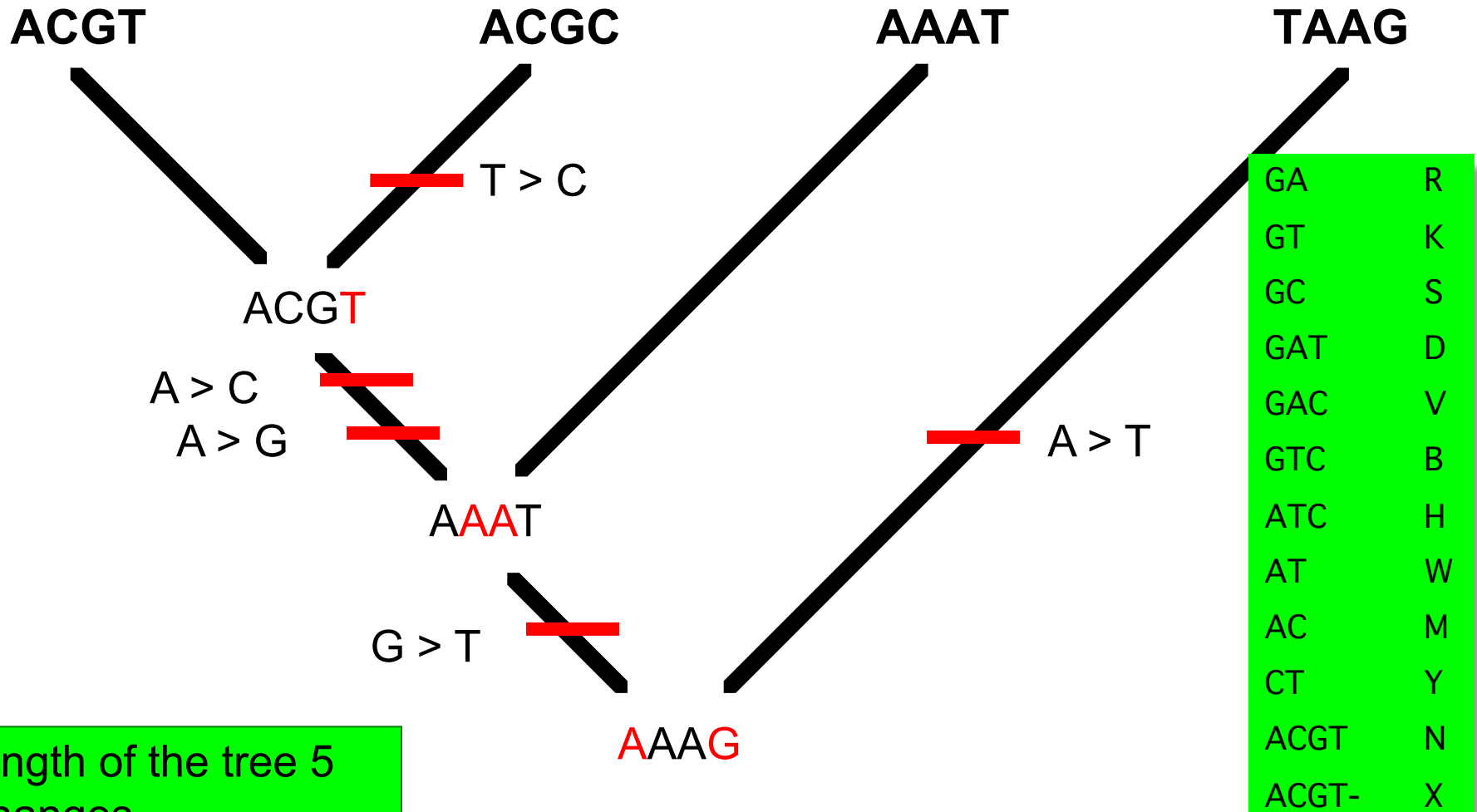
ATTENTION!

position = character  
nucleotide = ch. state

Each nucleotide treated INDEPENDENTLY

# Direct optimization (DO)

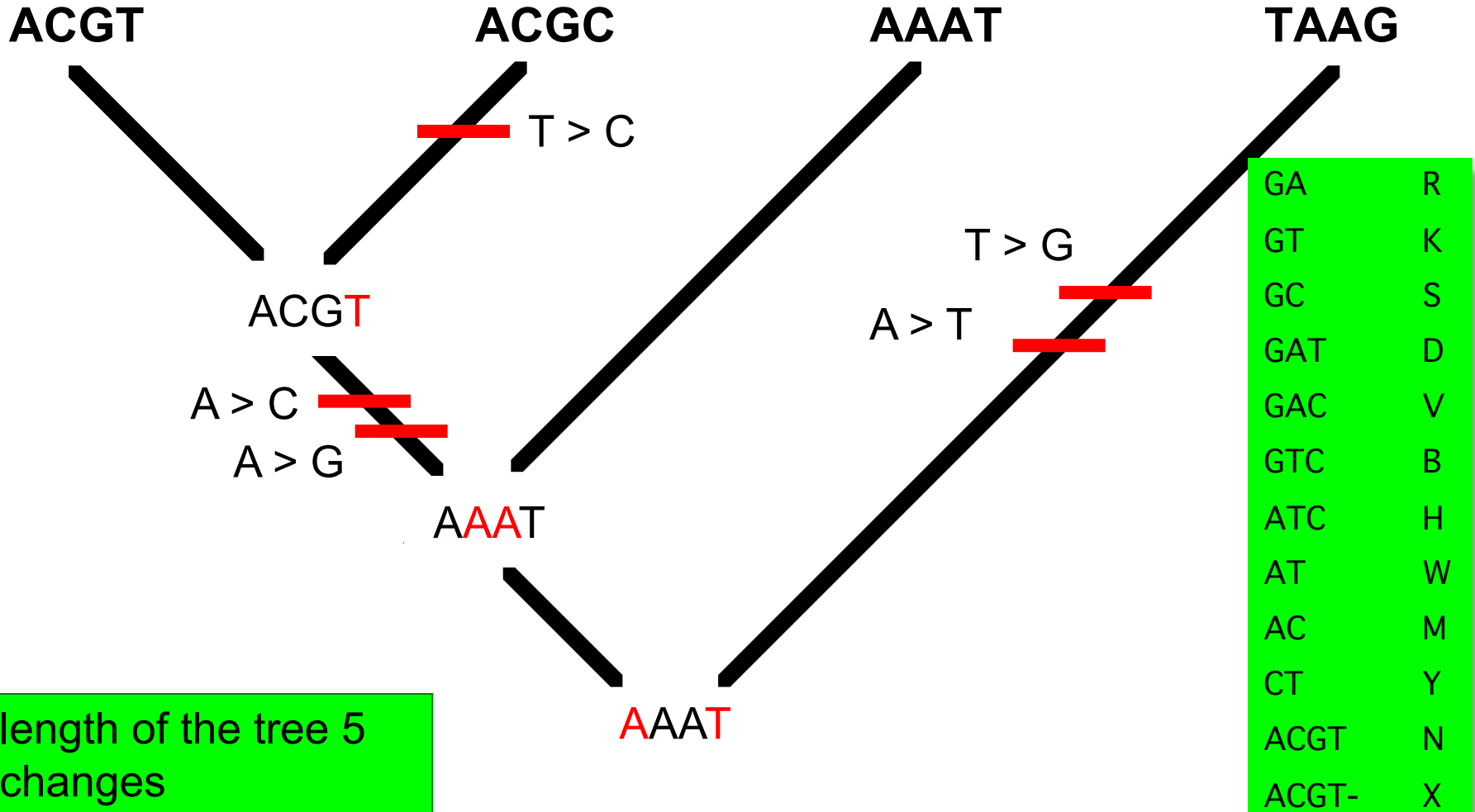
AACGGTTAAGGTACGGAGAATTAGGCAACCCTAGGATGCGATGCGCAGAGTTAGGTACGGAGAATTAGGCAACCCTAGGAT  
 AACGGTTAAGGTACGGAGAATTAGGCAACCCTAGGATGCGATGCGCAGAGTTAGGTACGGAGAATTAGGCAACCCTAGGATGCG



length of the tree 5 changes

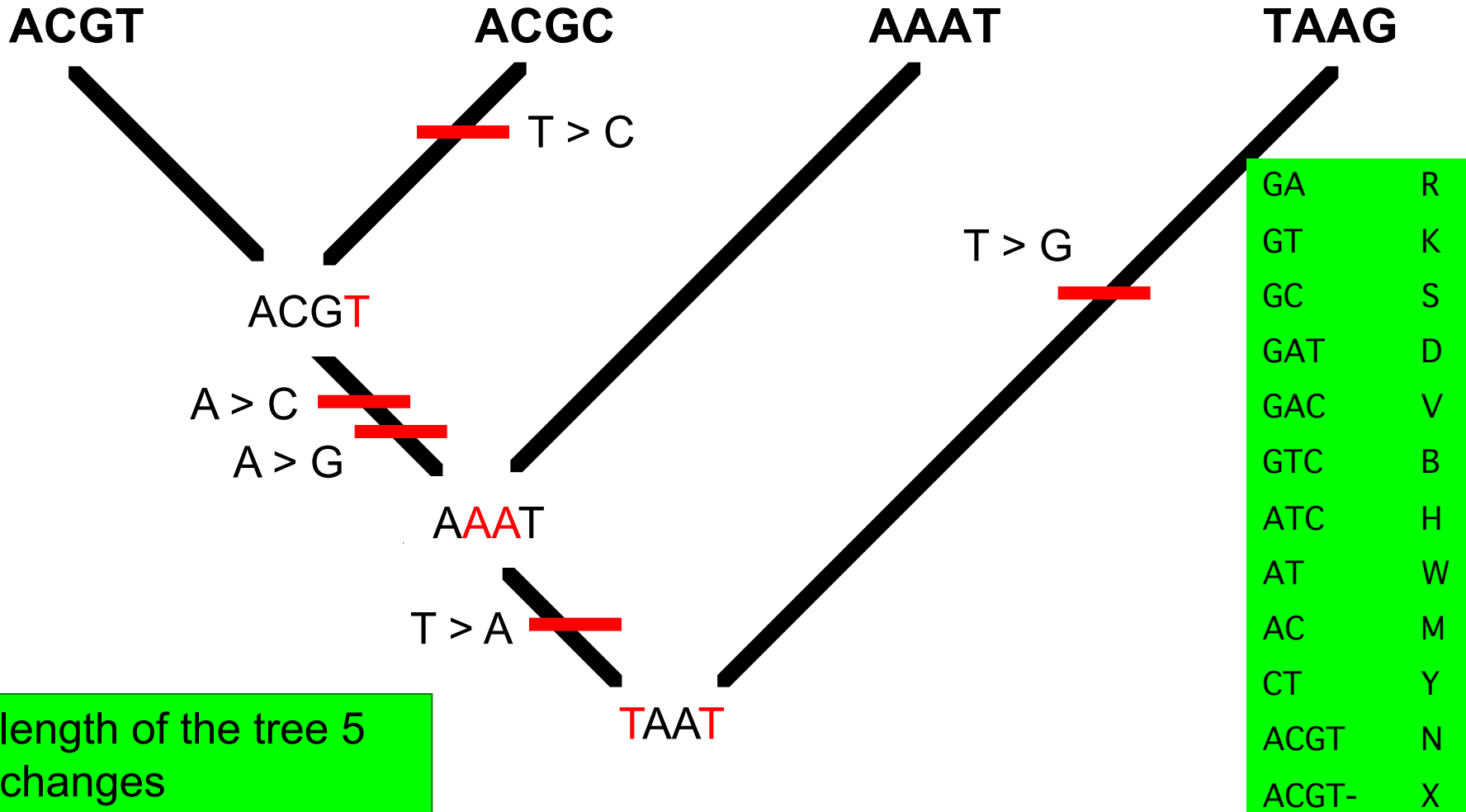
# Direct optimization (DO)

AACGGTTTAAGGTACGGAGAATTAGGCAACCCCTAGGATGCGATGCGCAGAGTTAGGTACGGAGAATTAGGCAACCCGGTAGGAT  
 AACGGTTTAAGGTACGGAGAATTAGGCAACCCCTAGGATGCGATGCGCAGAGTTAGGTACGGAGAATTAGGCAACCCCTAGGATGCG



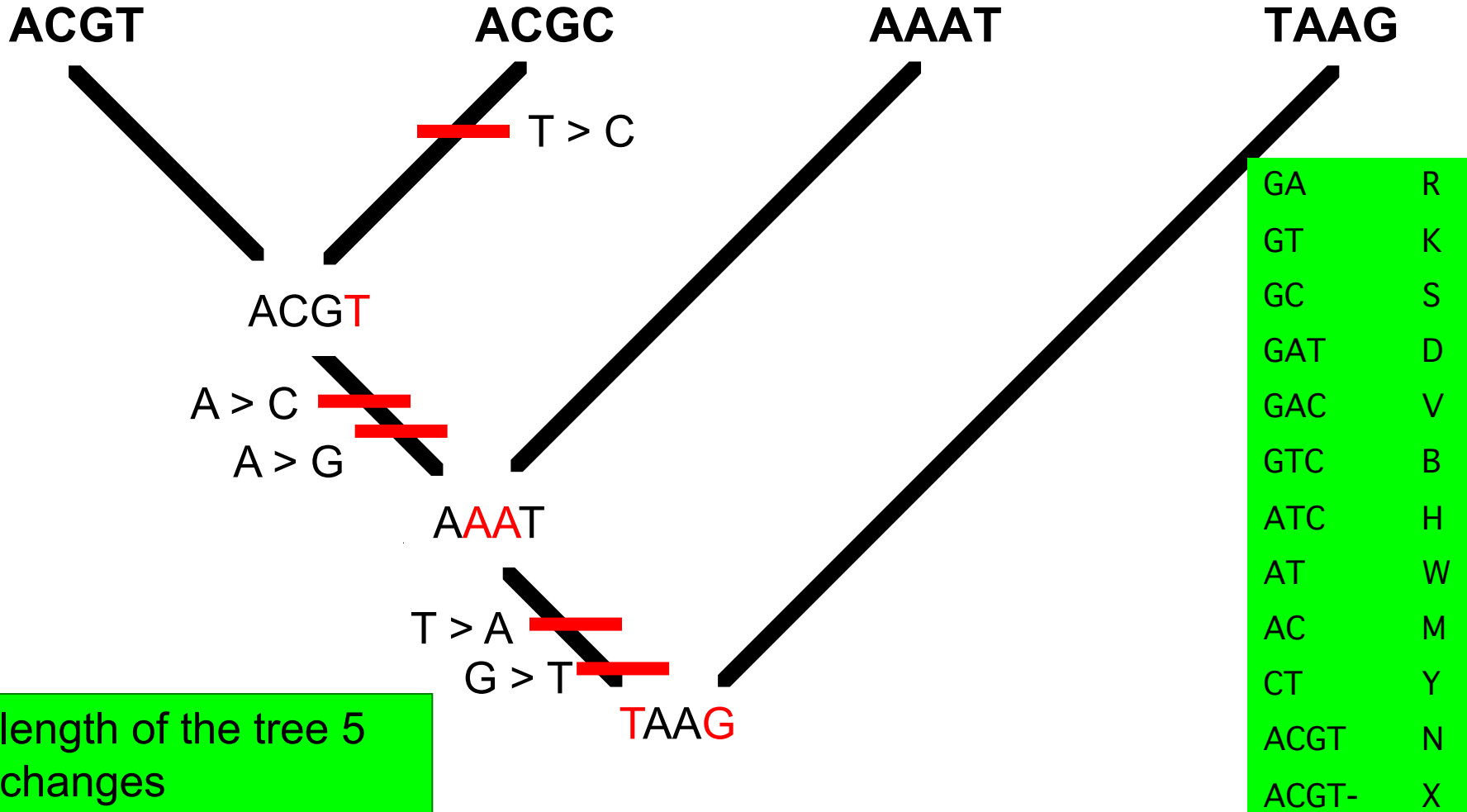
# Direct optimization (DO)

AACGGTTAAGGTACGGAGAATTAGGCAACCCTAGGATGCGATGCGCAGAGTTAGGTACGGAGAATTAGGCAACCCTAGGAT  
 AACGGTTAAGGTACGGAGAATTAGGCAACCCTAGGATGCGATGCGCAGAGTTAGGTACGGAGAATTAGGCAACCCTAGGATGCG



# Direct optimization (DO)

AACGGTTAAGGTACGGAGAATTAGGCAACCCTAGGATGCGATGCGCAGAGTTAGGTACGGAGAATTAGGCAACCCTAGGAT  
 AACGGTTAAGGTACGGAGAATTAGGCAACCCTAGGATGCGATGCGCAGAGTTAGGTACGGAGAATTAGGCAACCCTAGGATGCG



# SUMMARY

optimization has to be used in order to find shortest tree &  
to find character states for internal nodes

MULTIPLE equally parsimonious reconstructions are possible

affects WHERE changes are posited

pay attention to how different reconstructions affect  
interpretation of the results obtained

alignments are **TOPOLOGY specific**