

+

+



✓



✓

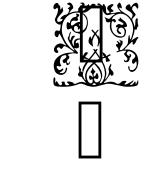
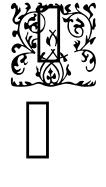


+

+



v



Y



Y



Y

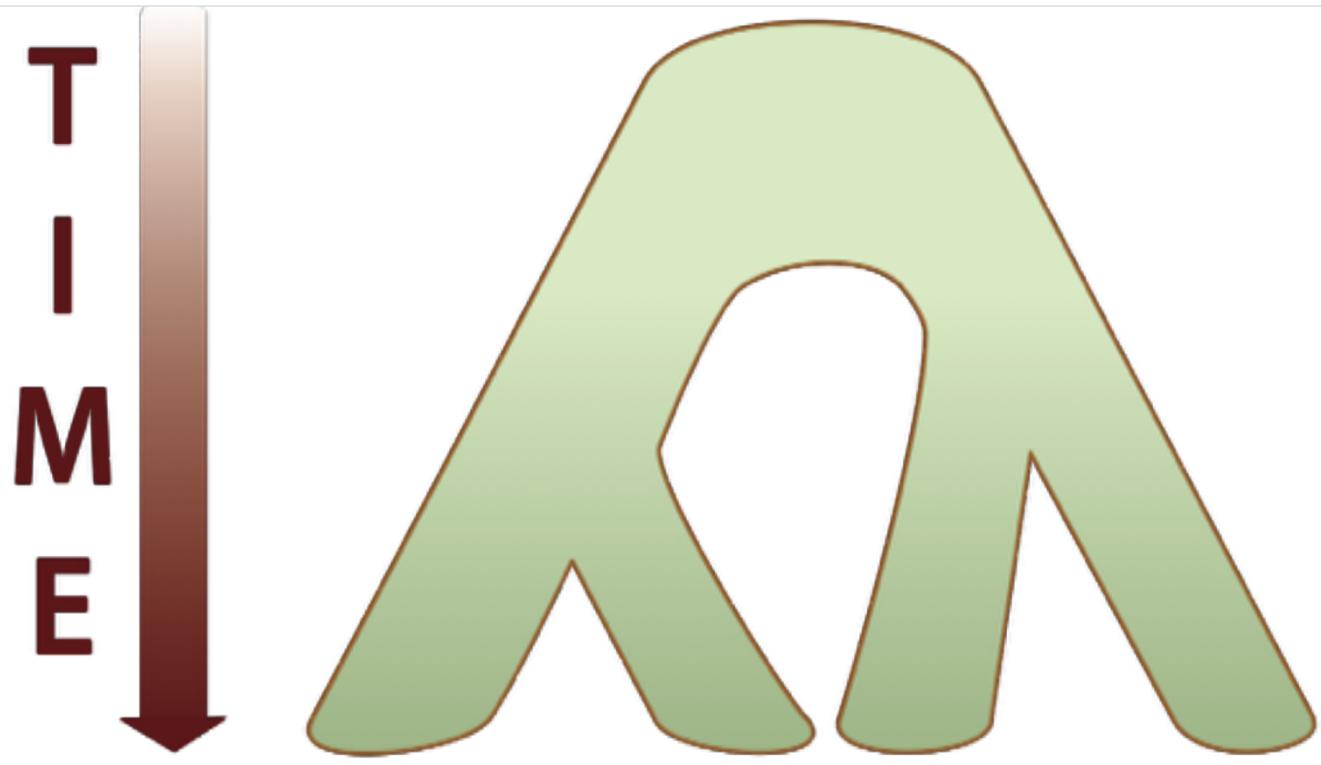


Y



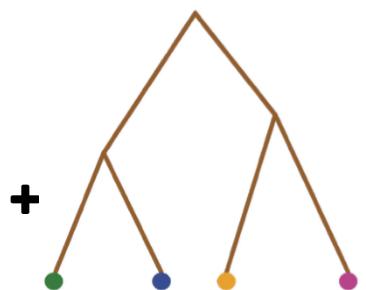
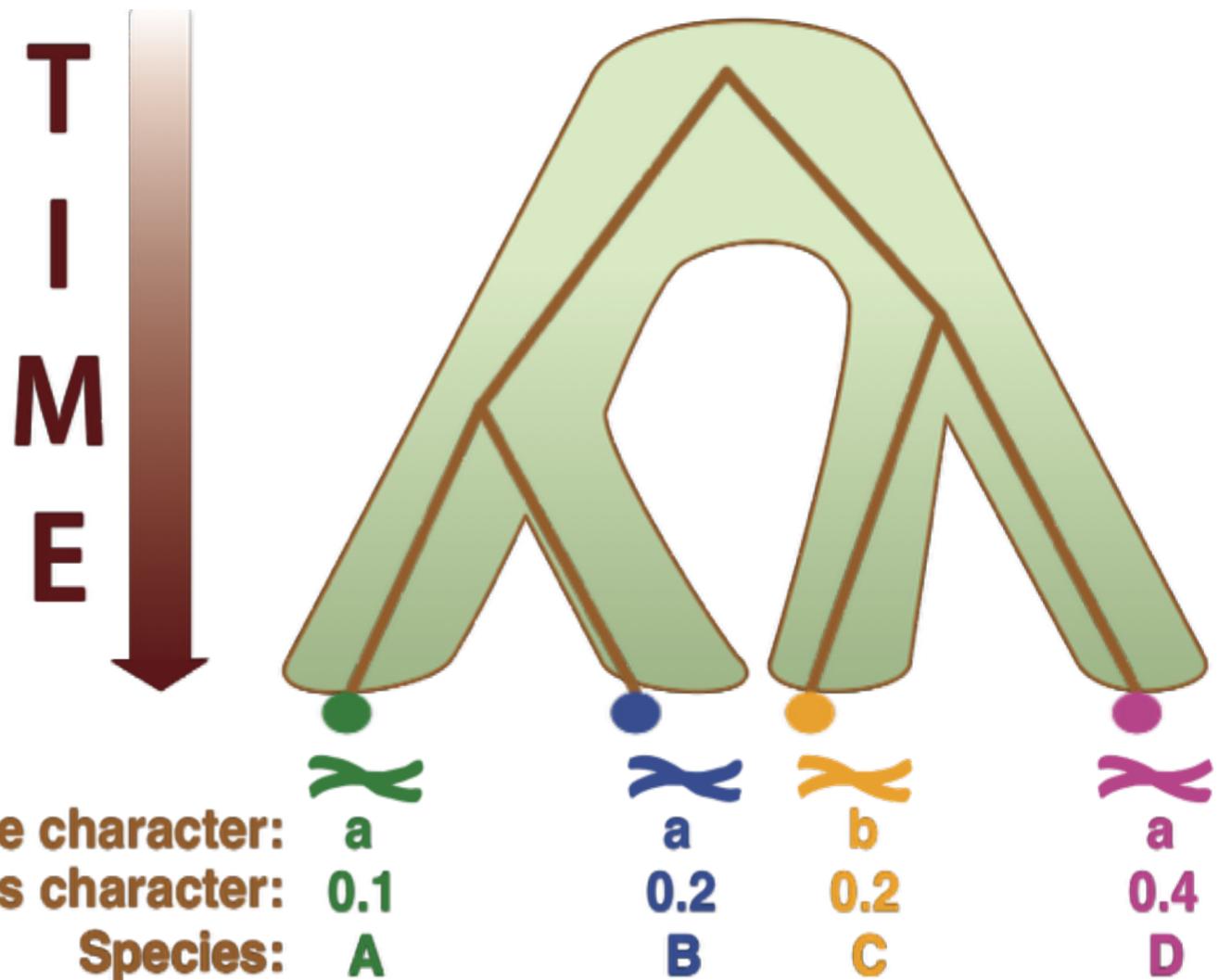
+

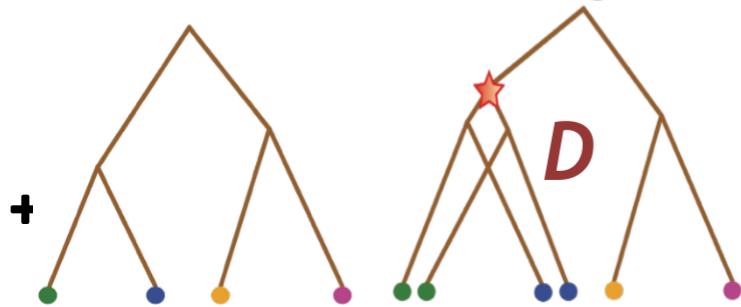
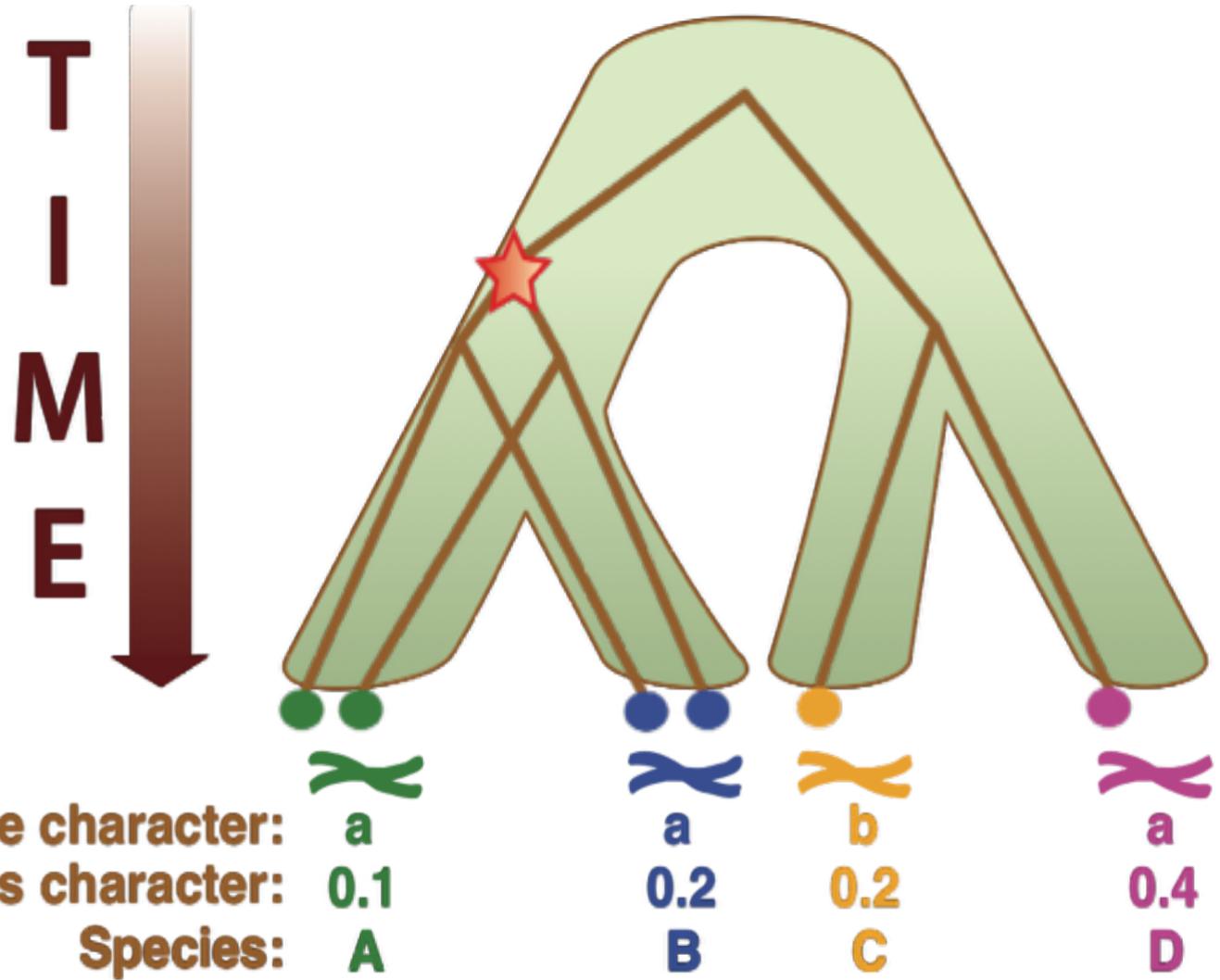
+

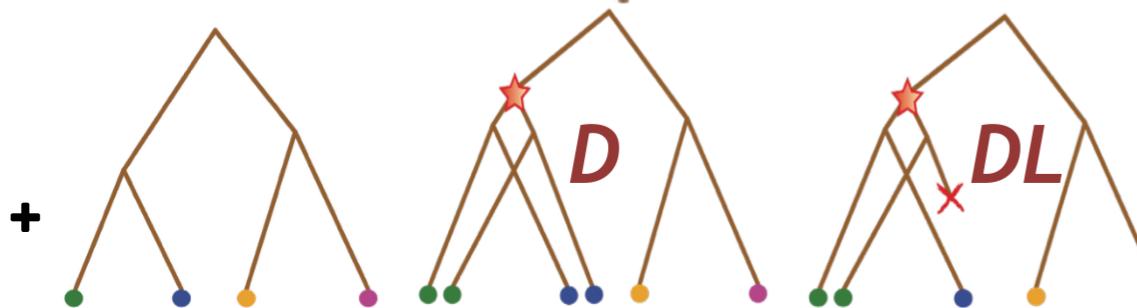
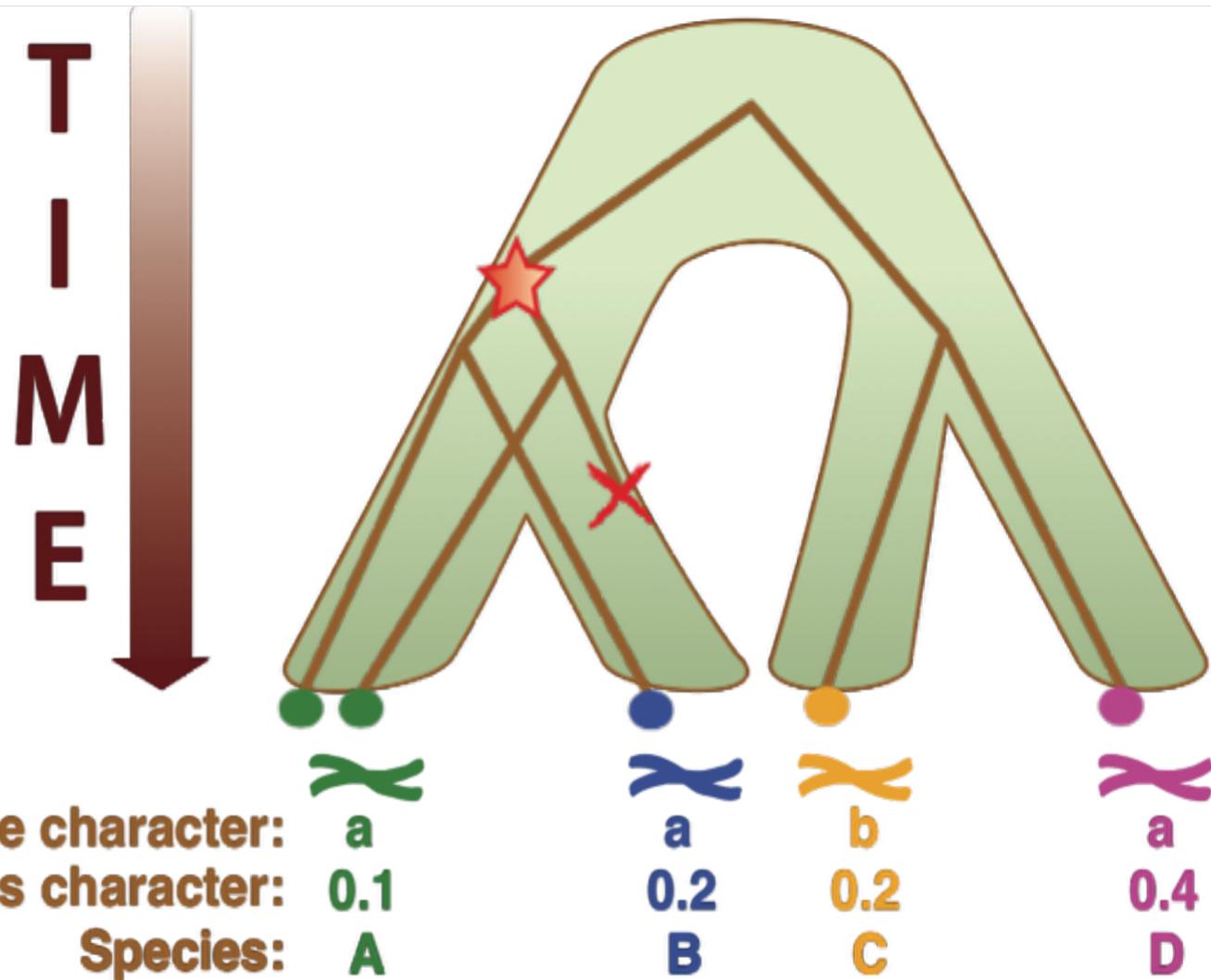


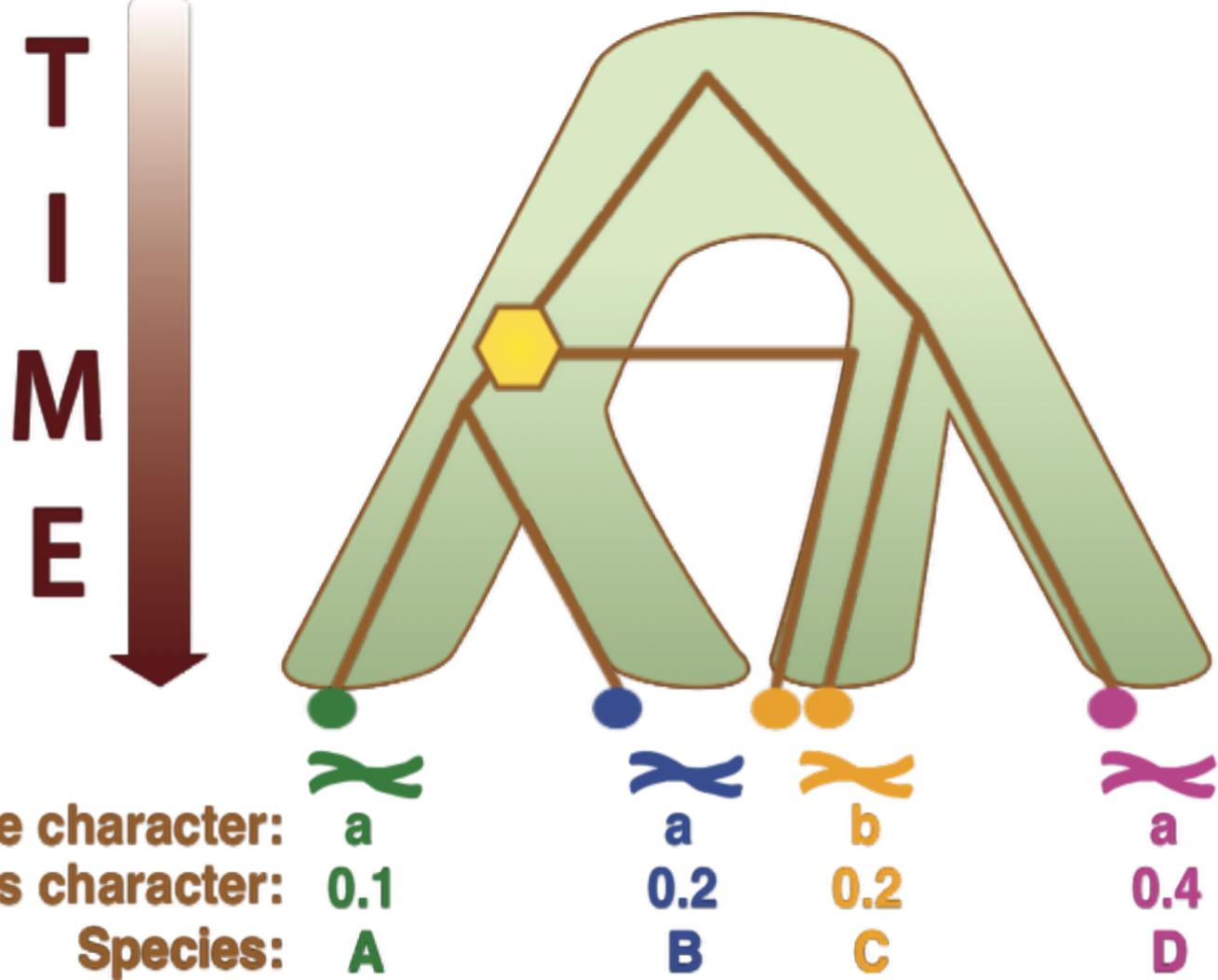
Discrete character:		a		
Continuous character:	0.1			a
Species:	A		B	C

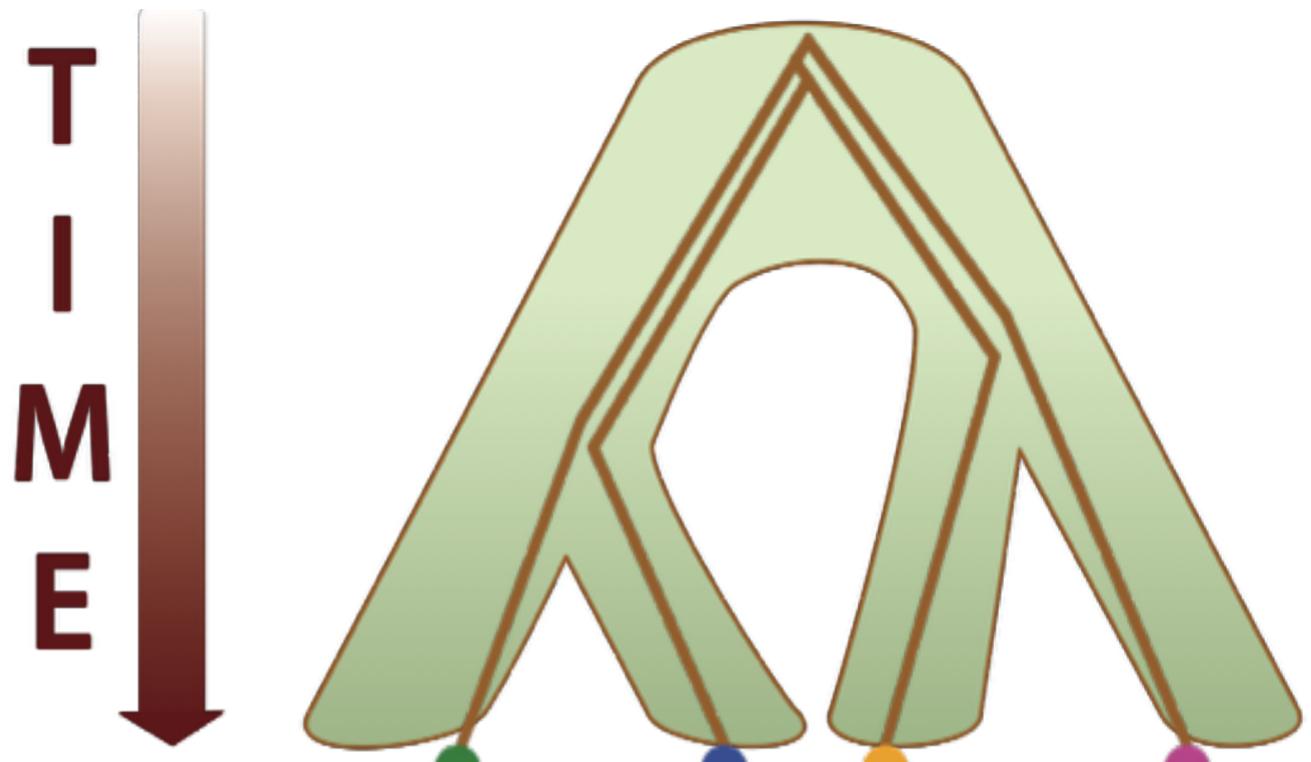
				b
			0.2	0.2
			D	











Discrete character:

Continuous character:

Species:

a

0.1

A

a

0.2

B

b

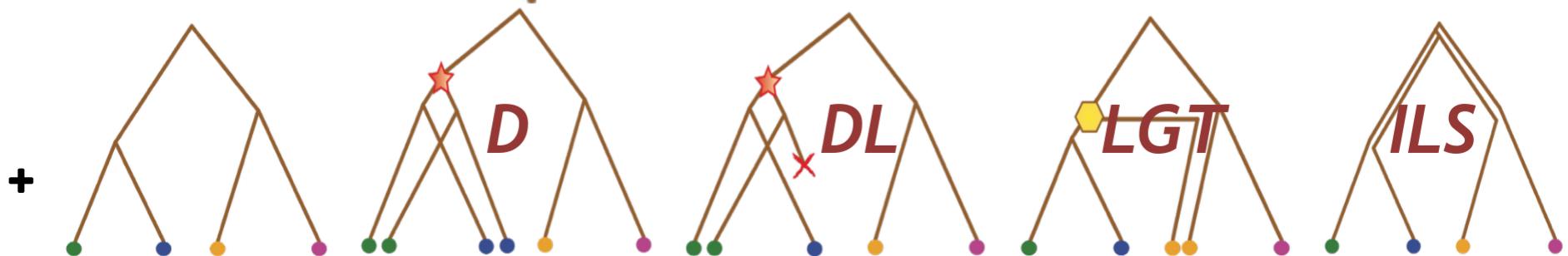
0.2

C

a

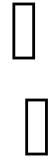
0.4

D





v



Y



Y



Y



Y

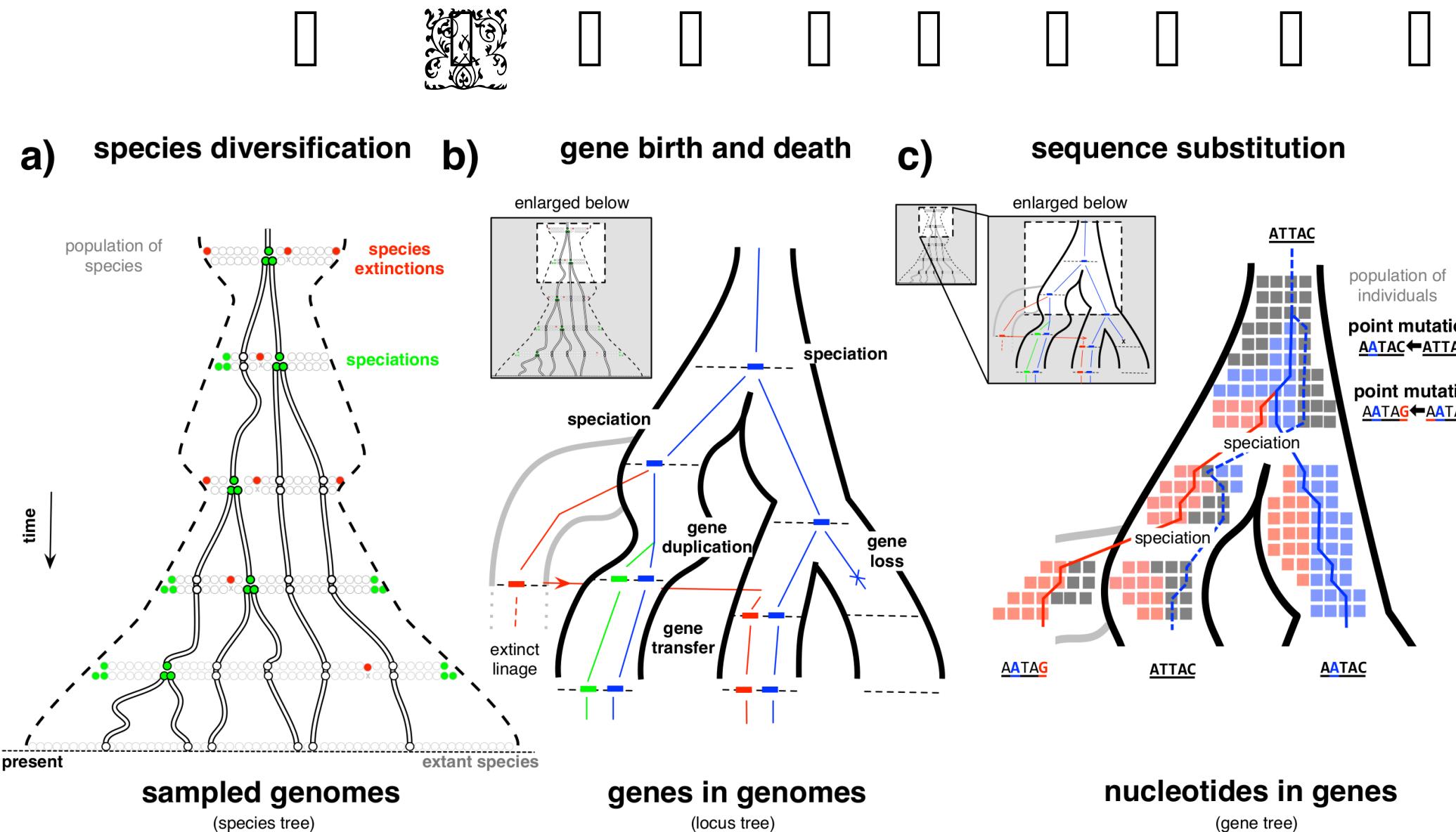


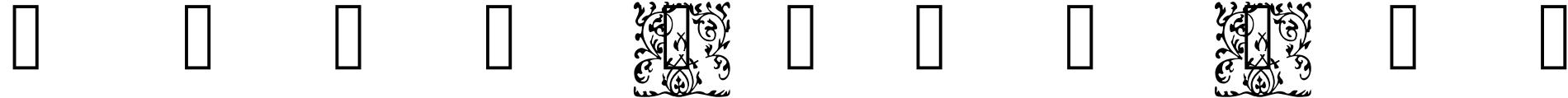
v



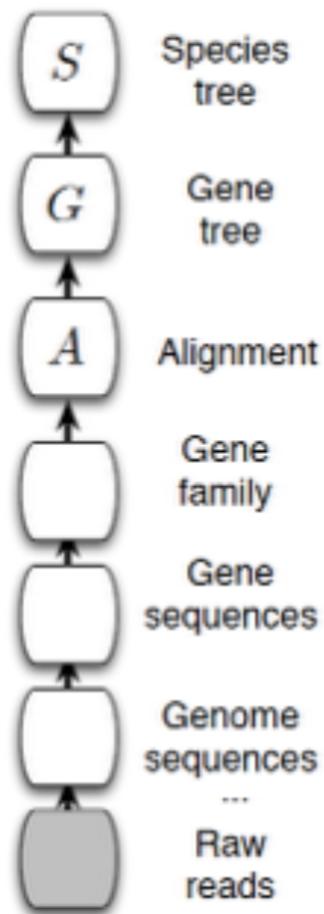
+

+

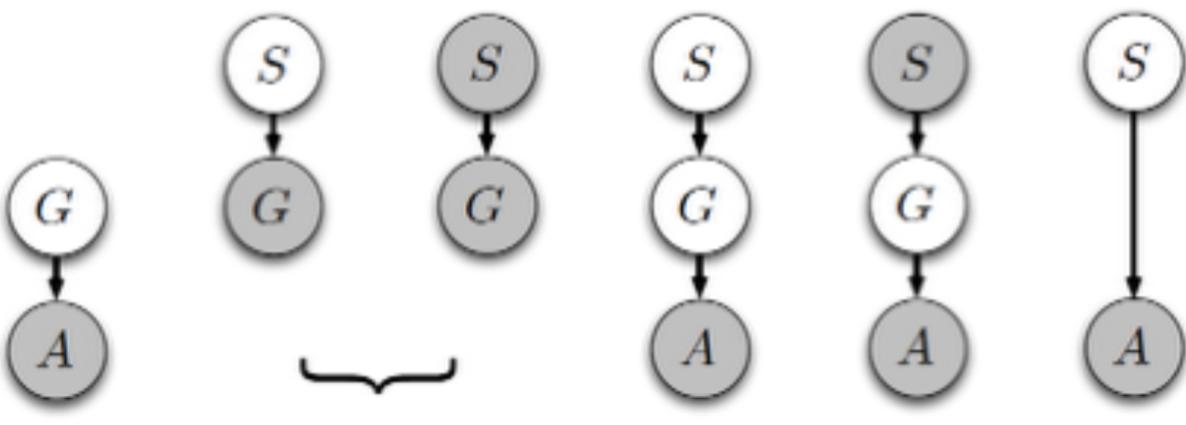




Phylogenomics inference pipeline



Gene tree-species tree models published in the literature



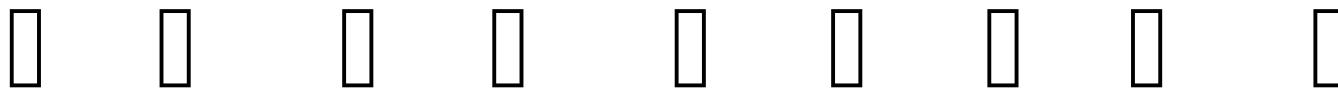
$P(A|G)$
PhyML
 [Guindon & Gascuel, 2003];
RAXML
 [Stamatakis et al., 2005];
MrBayes
 [Ronquist et al., 2012];
BEAST
 [Drummond et al., 2012].

$P(G|S)$
BEST
 [Edwards et al., 2007];
MP-EST
 [Liu et al., 2010];
ODT
 [Szöllősi et al., 2012];

$P(A|G, S)$
***BEAST**
 [Heled & Drummond, 2008];
PHYLDOG
 [Boussau et al., 2013];
Prime-DLRS
 [Sjöstrand et al., 2012];
exODT
 [Szöllősi et al., 2013a];

$P(A|S)$
Prime-GSR
 [Akerborg et al., 2009];
IMa2
 [Hey, 2011];
SNAP
 [Bryant et al., 2012];
ALE
 [Szöllősi et al., 2013b];
POMO
 [De Maio et al., 2013]





This decorative horizontal border consists of a repeating pattern of small, empty rectangular boxes. It is punctuated by two larger, ornate floral brackets, one on the left and one on the right, which are symmetrical in design.

+



✓ □

✓ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □

✓ □

✓ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □

□ □ □ □ □ □ □ □ □ □

Y □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □

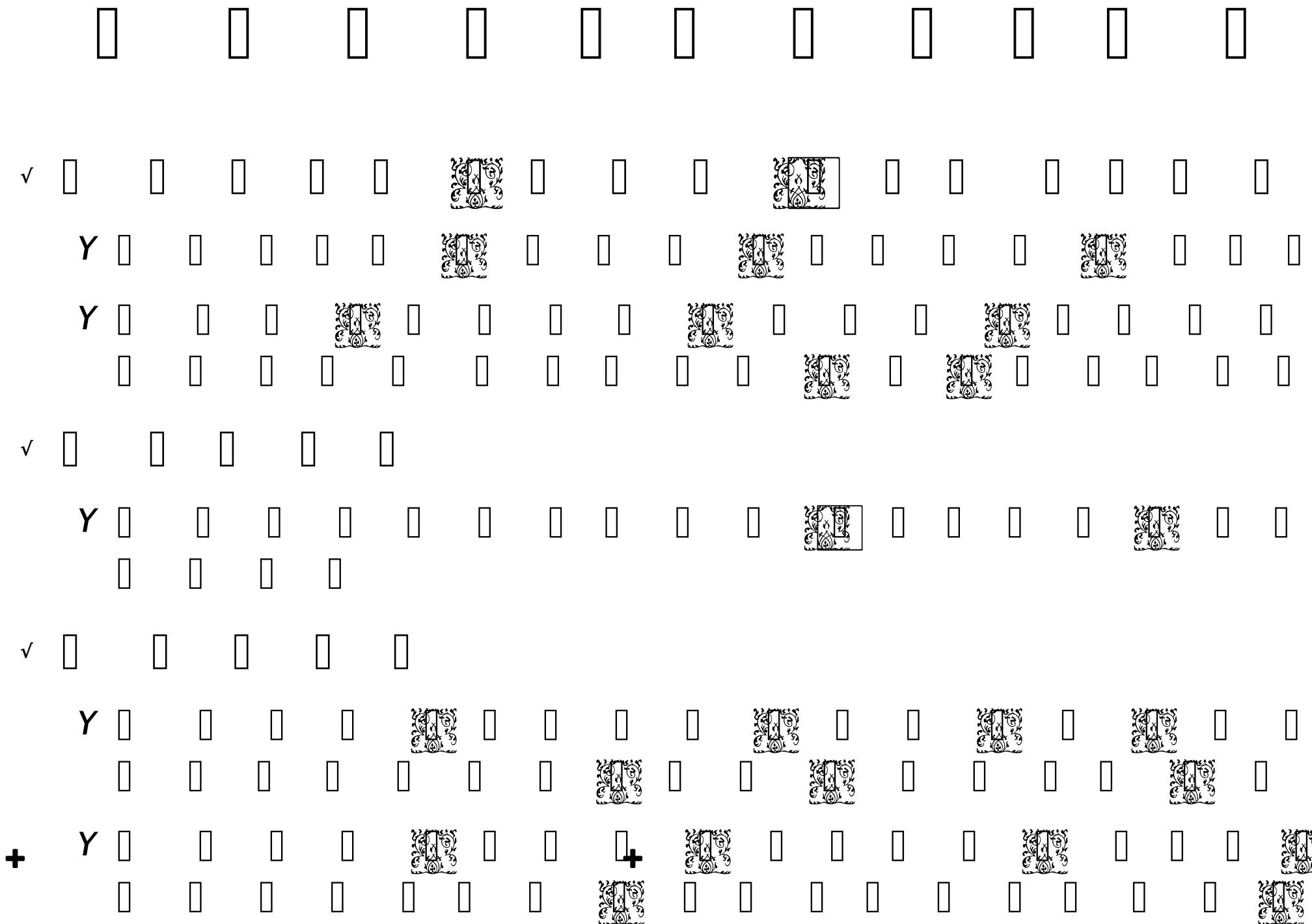
✓ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □

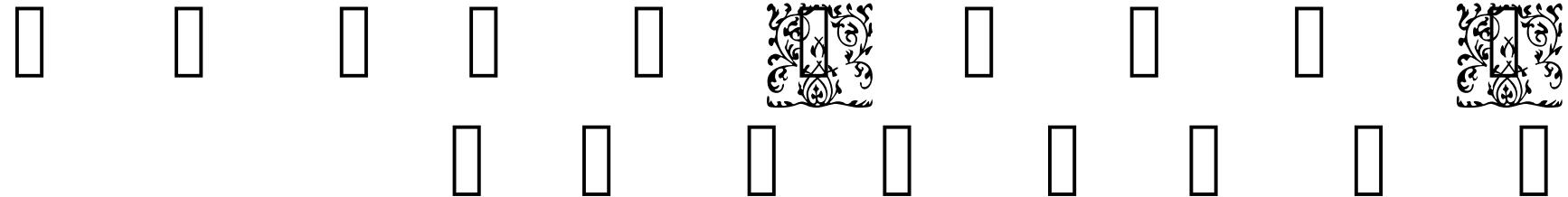
□ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □

□ □ □ □ □ □ □ □

+

+





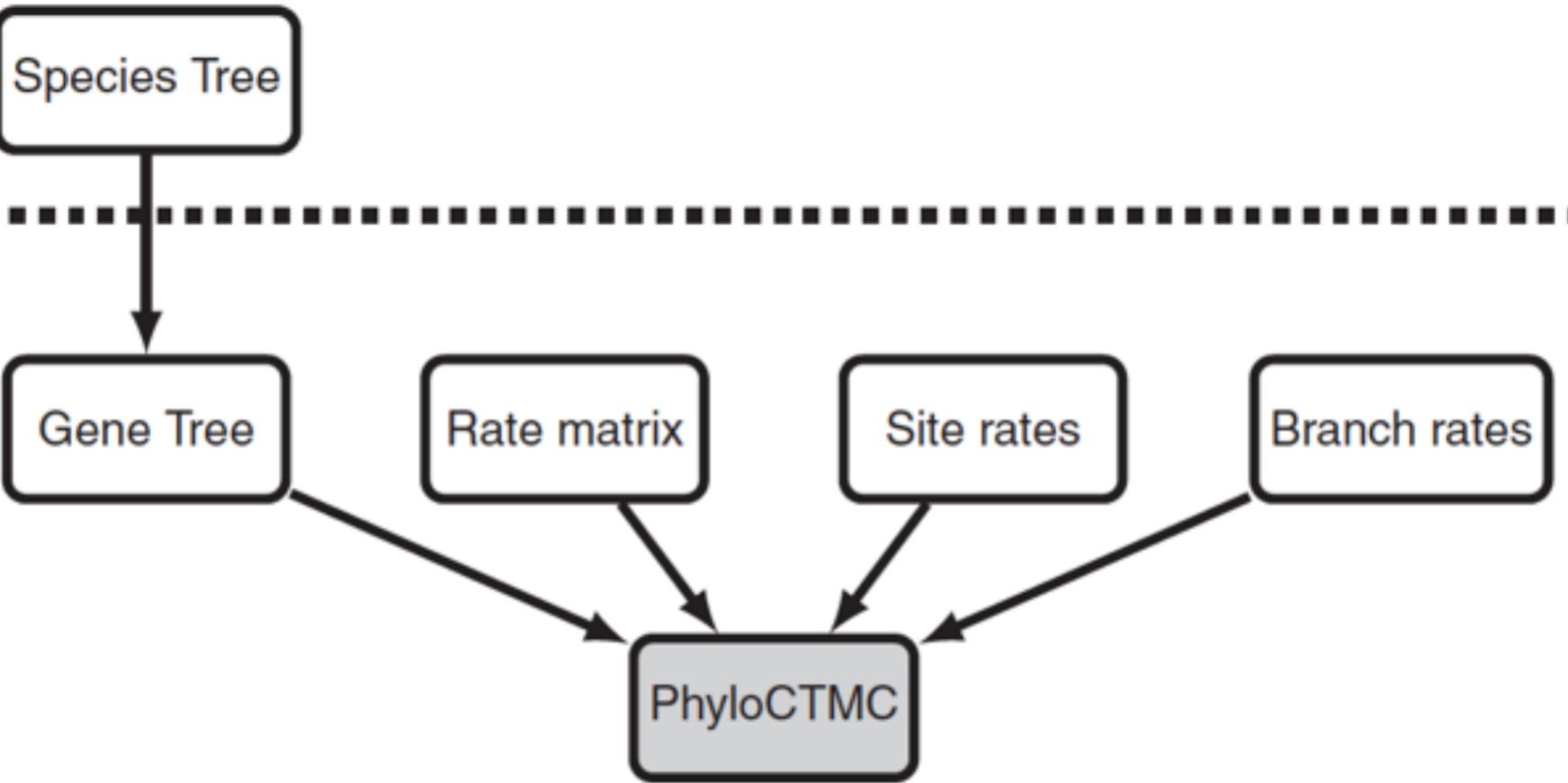
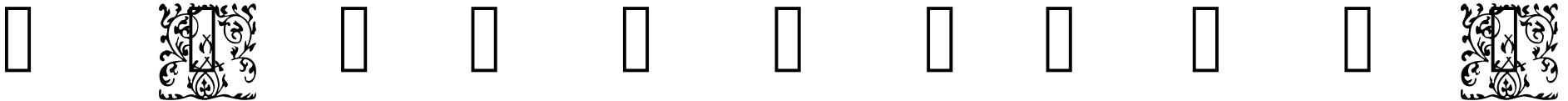
v □ □ □ □ □

Y □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □
□ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □

Y □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □
□ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □
□ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □
□ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □

+

+

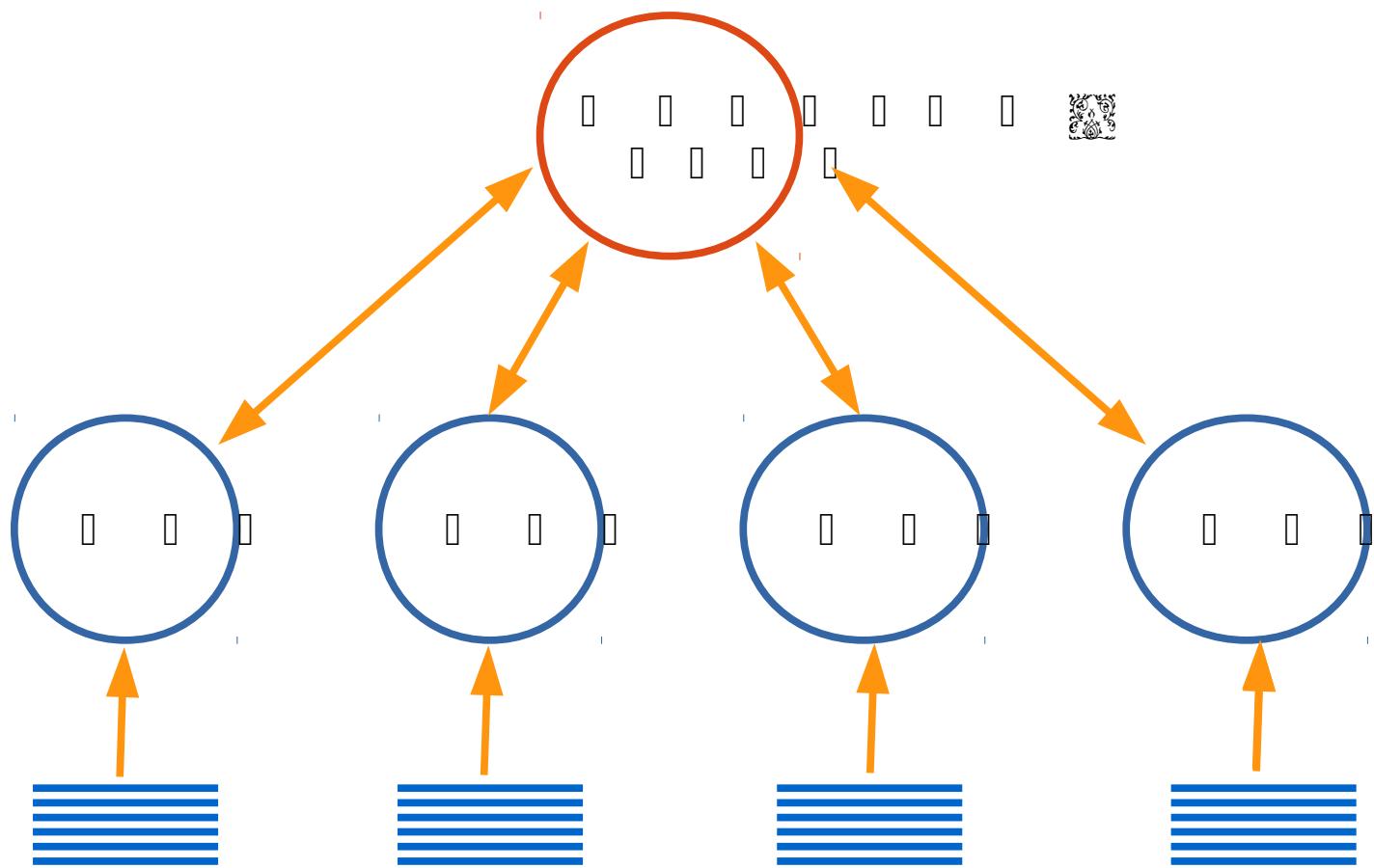
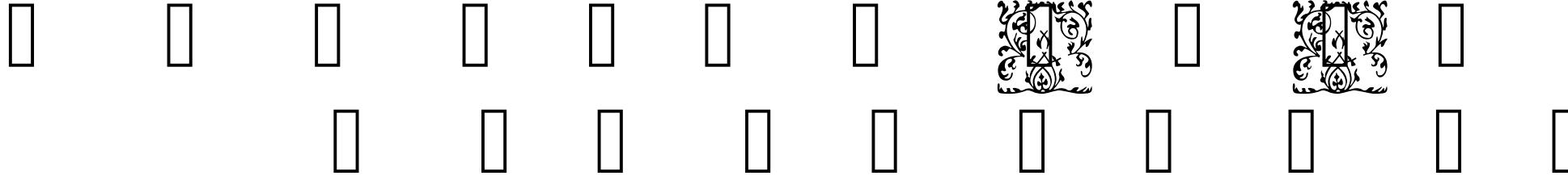


gene family $\in \{1, 2, \dots, N\}$

+

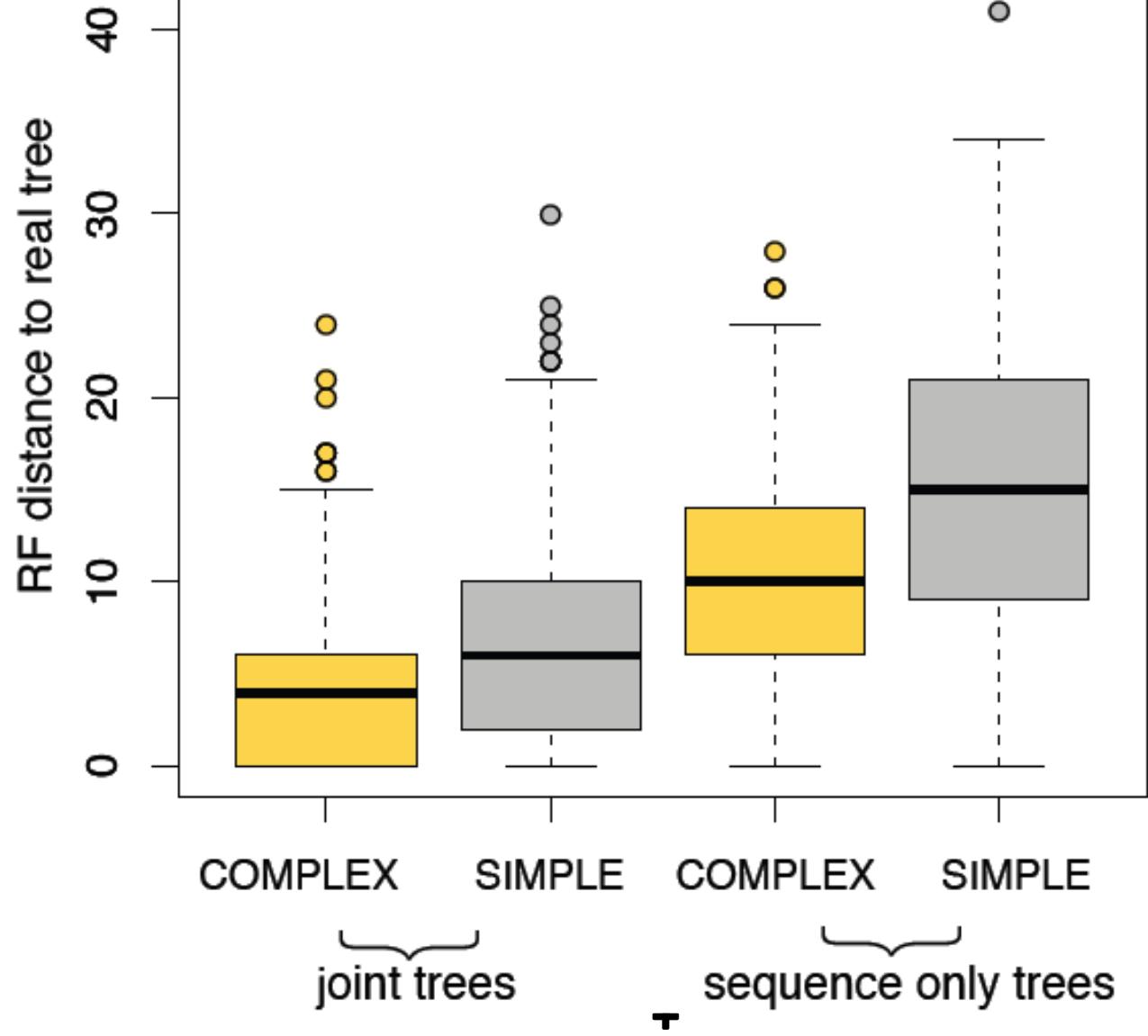
+



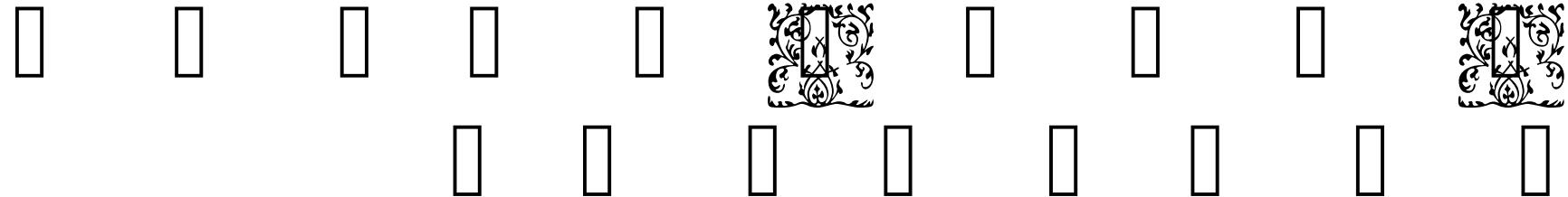


+

+



+



v □ □ □ □

Y □

Y □

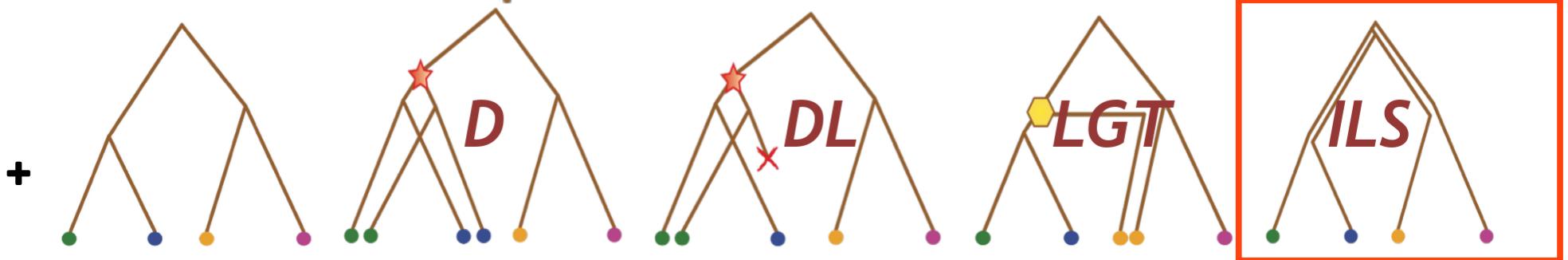
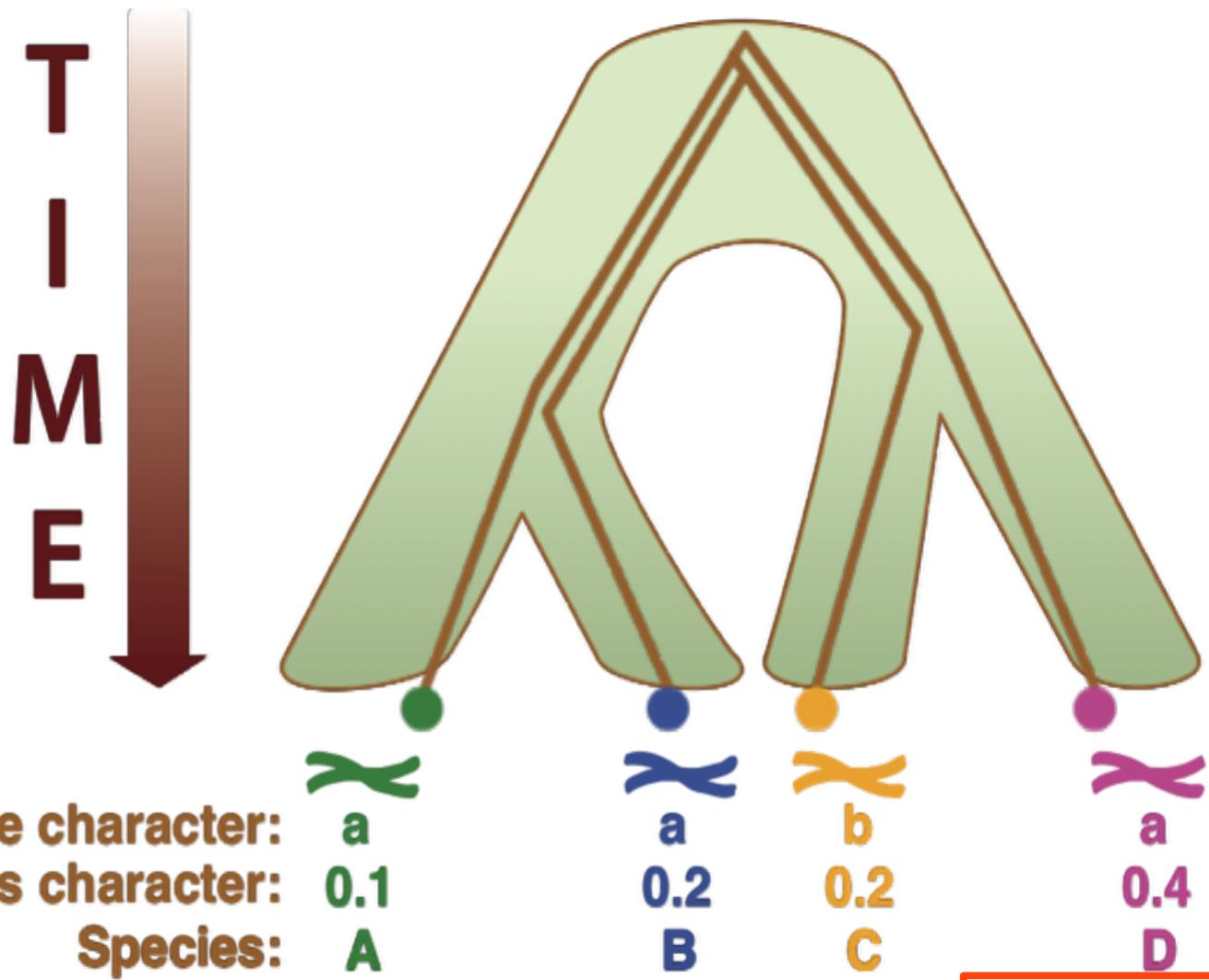
v □ □ □ □

Y □

Y □

+

+



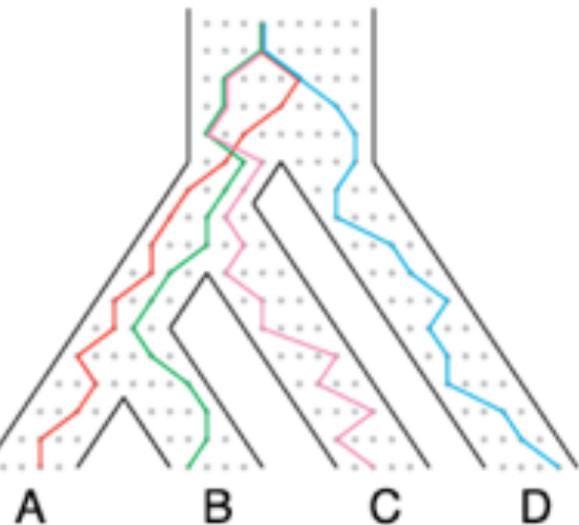
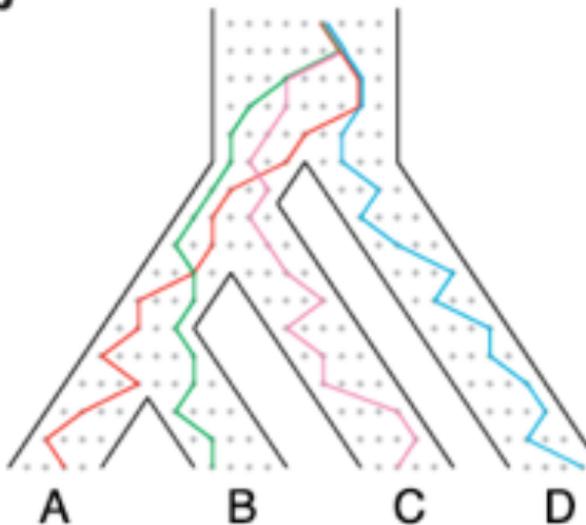
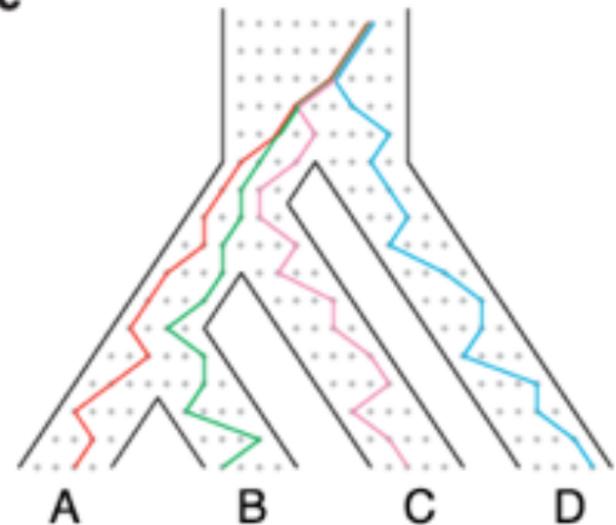


✓														
✓														
✓														
✓														
✓														
✓														
✓														
✓														
✓														
✓														

+

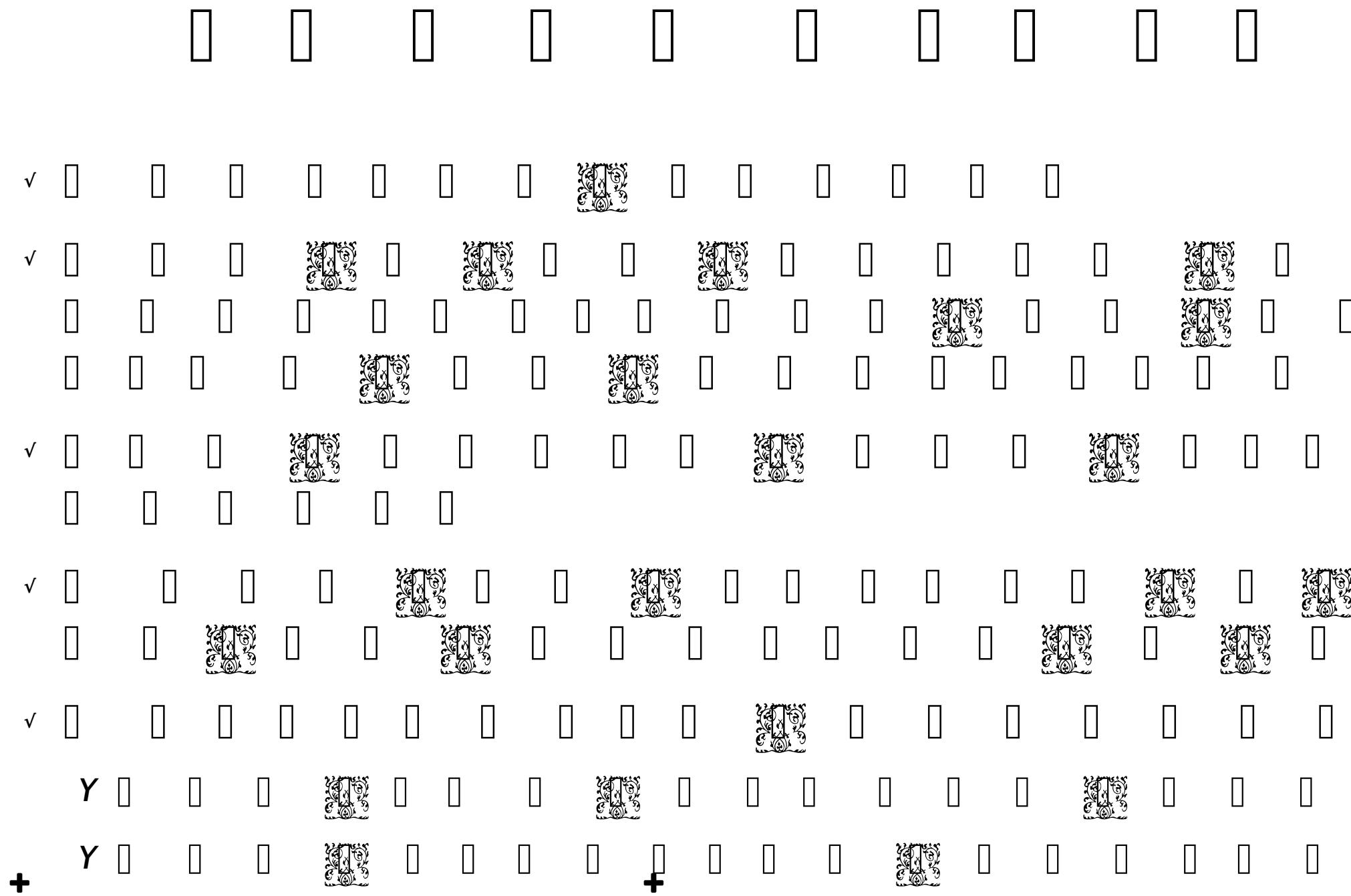
+

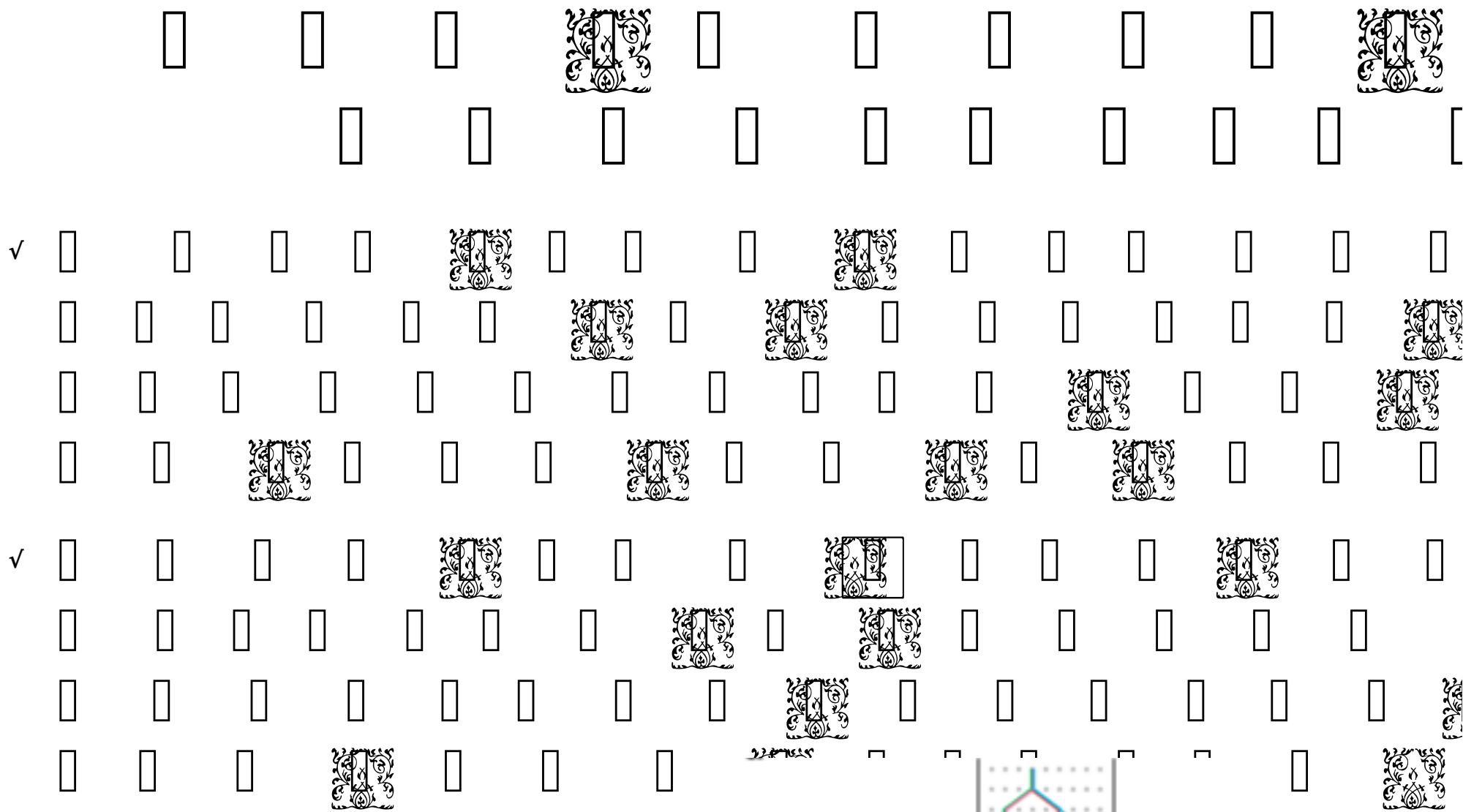


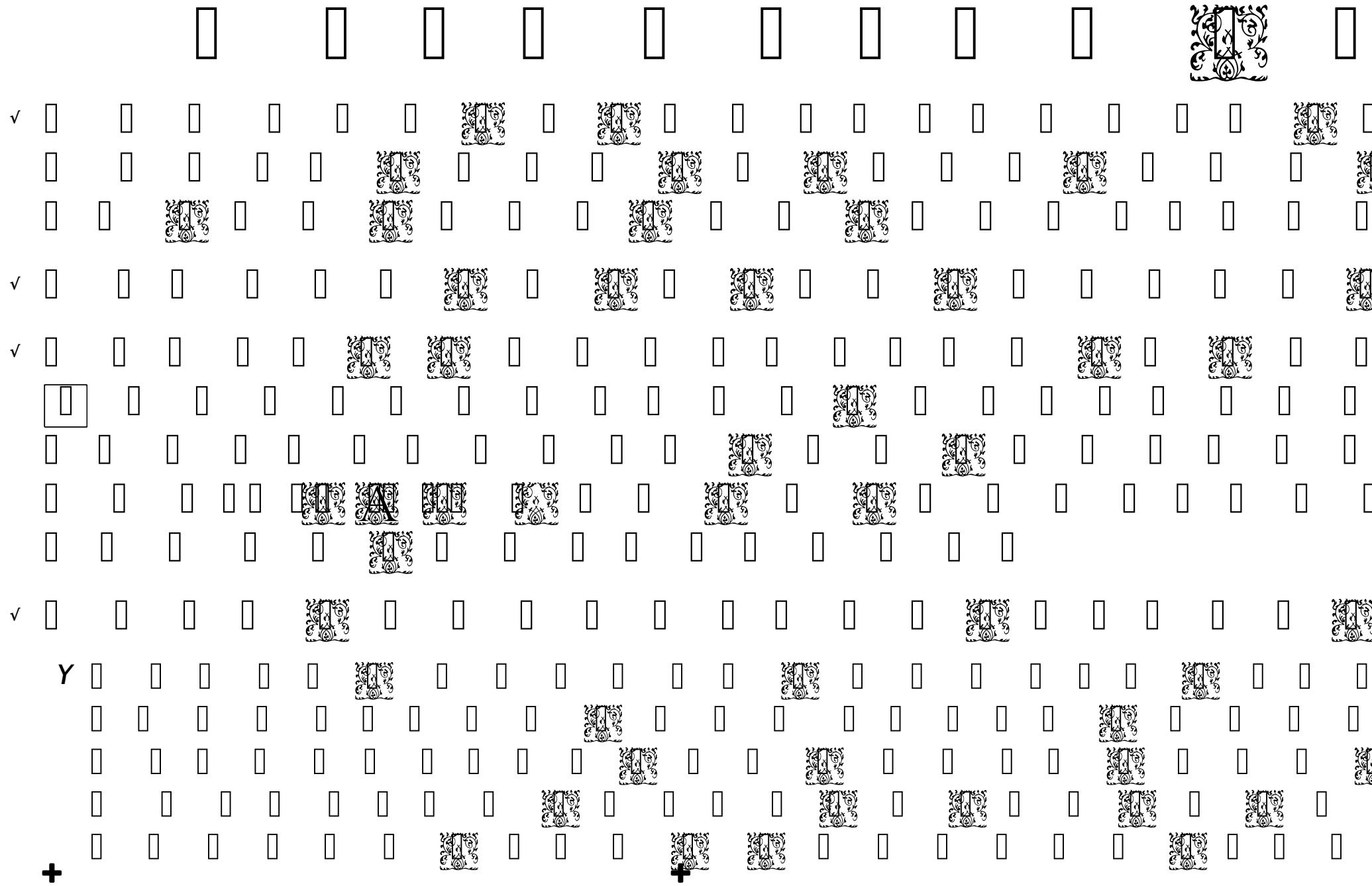
**a****b****c**

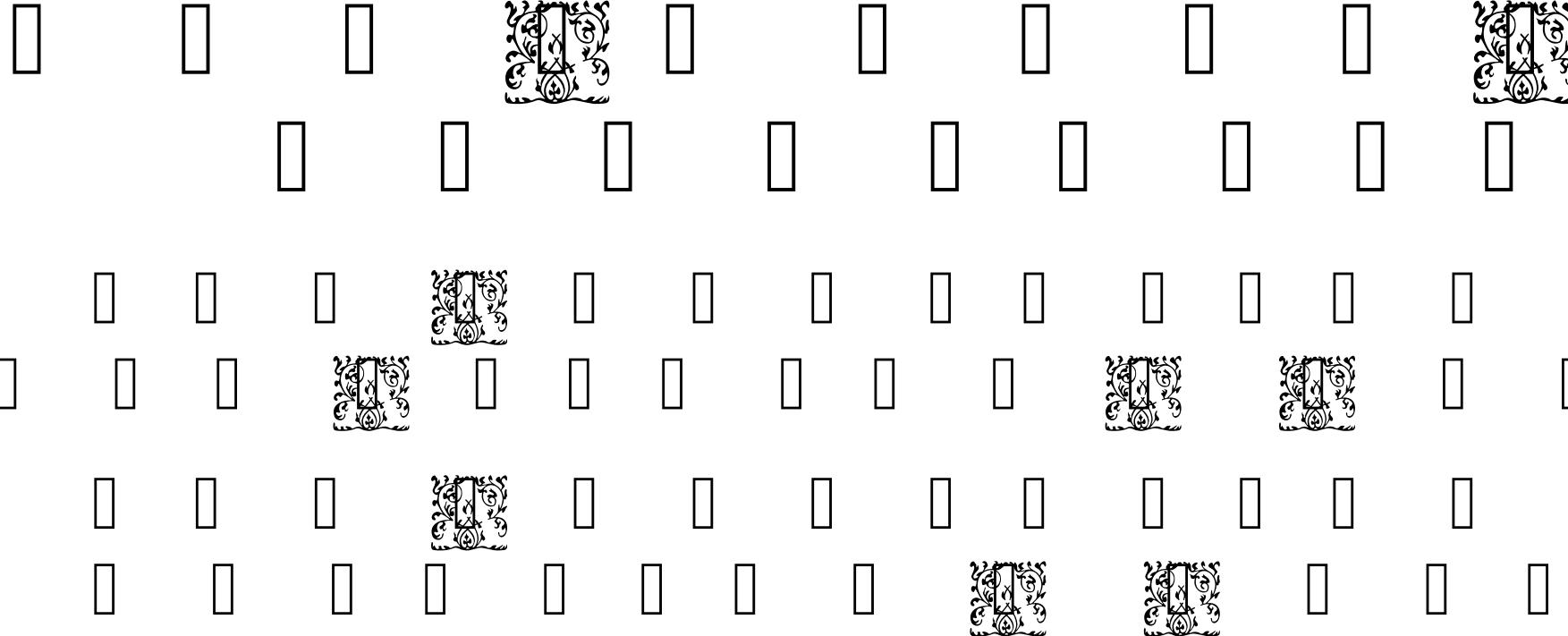
+

+



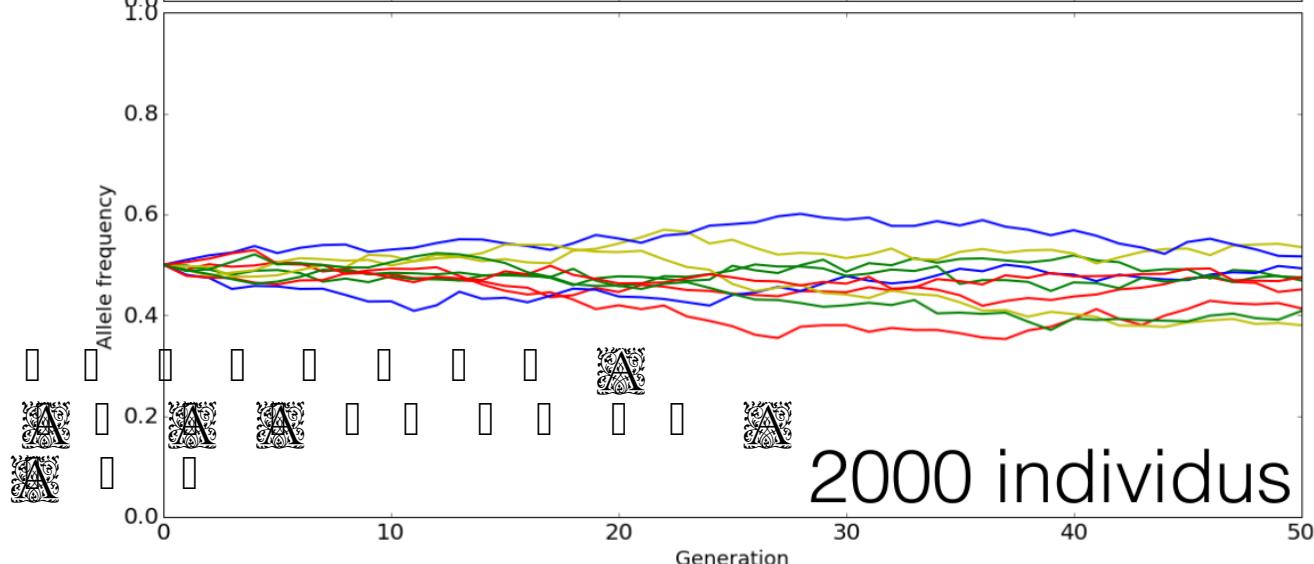
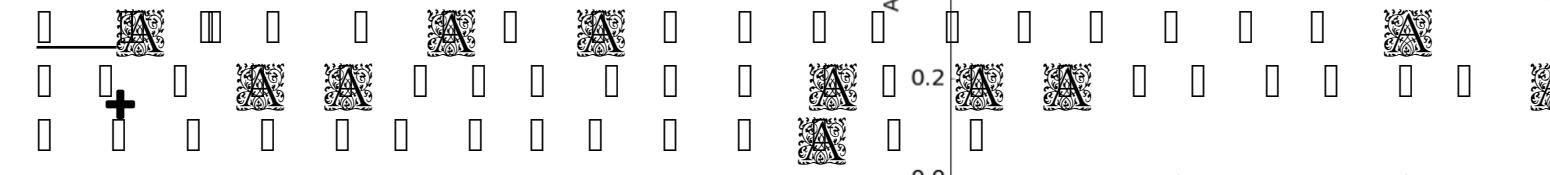
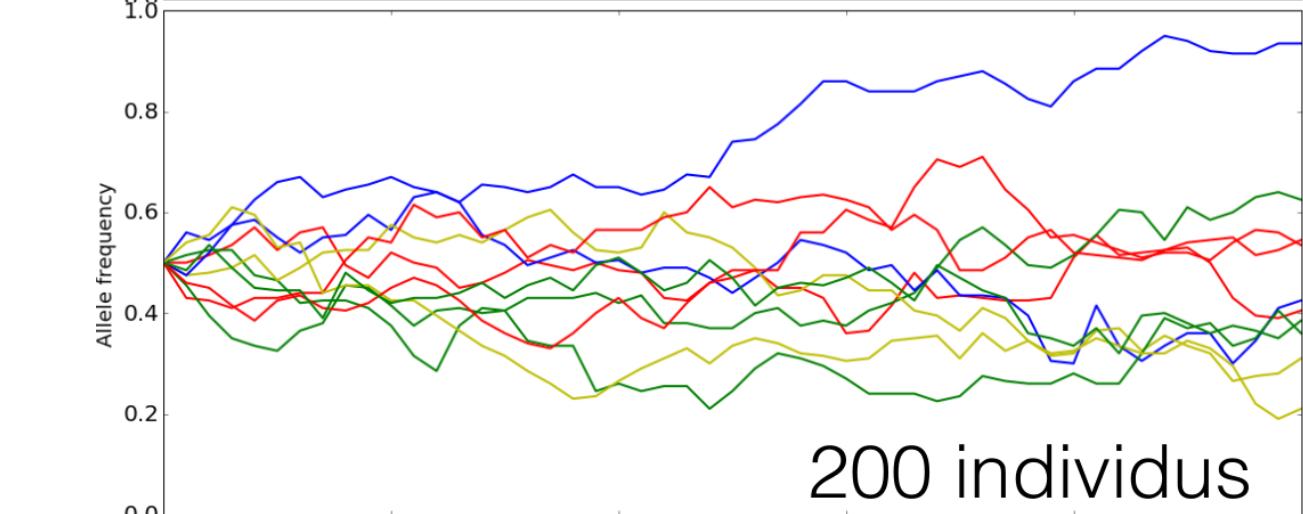
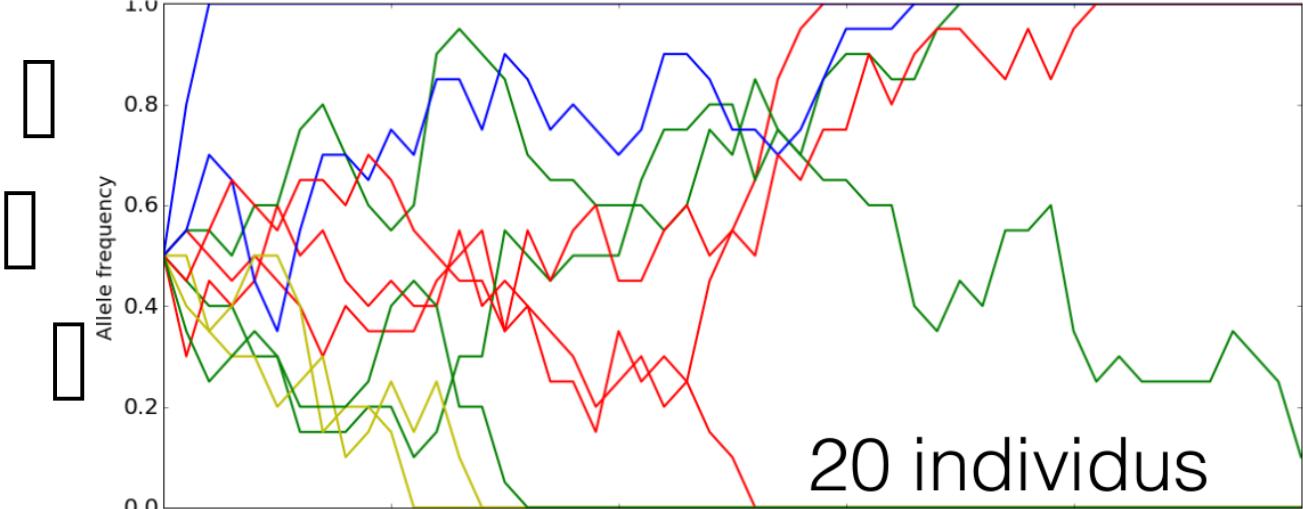
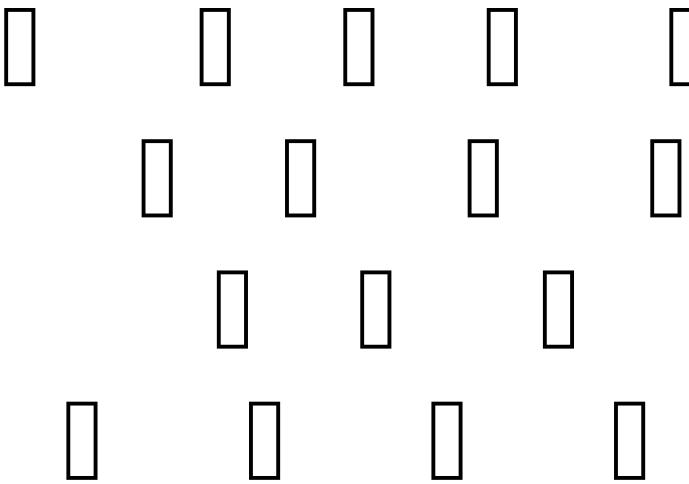


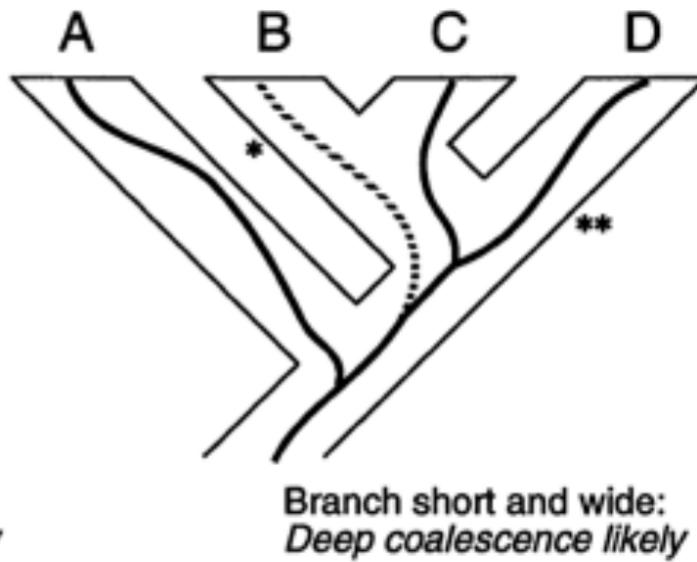
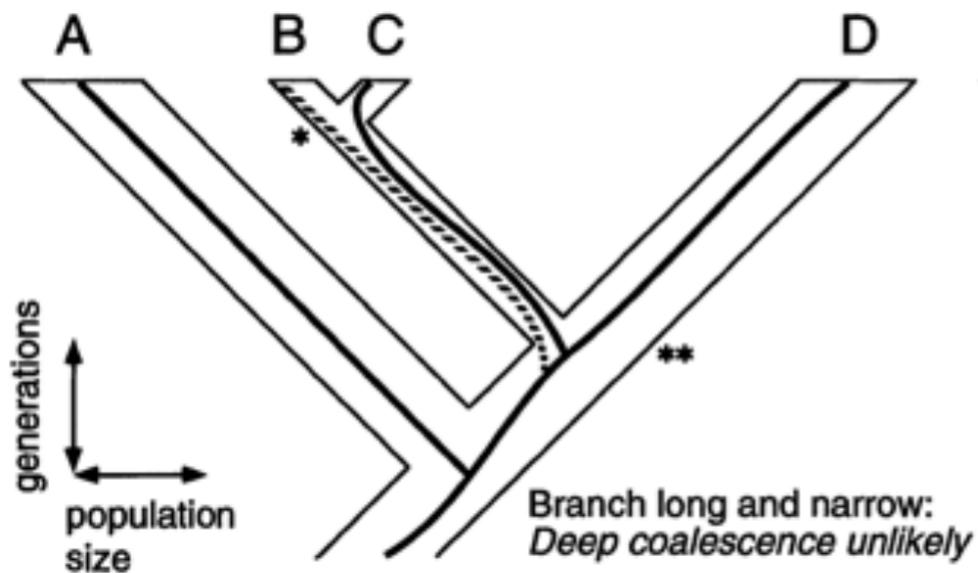
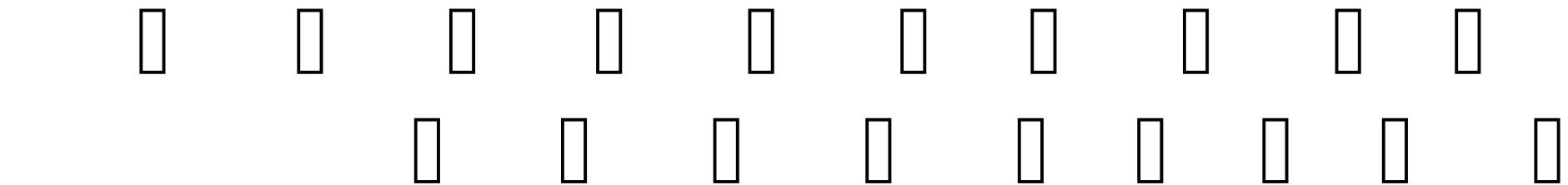




+

+



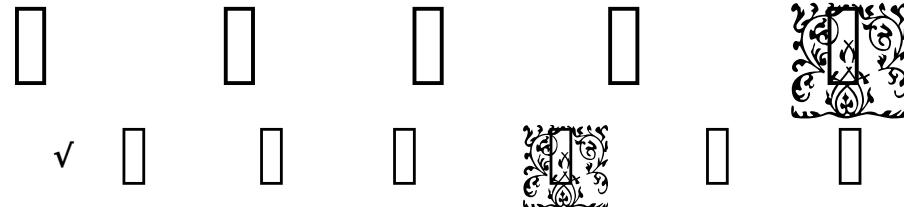


+

+



v



τ_{HCGO}

τ_{HCG}

τ_{HC}

Rannala and Yang, *Genetics* 2003

HCGO

$t_2^{(\text{HCGO})}$

$t_3^{(\text{HCGO})}$

HCG

d

$t_2^{(\text{HC})}$

$t_3^{(\text{HC})}$

H

a

$t_3^{(\text{H})}$

C

b

$t_2^{(\text{C})}$

H₁

H₂

H₃

C₁

C₂

G

O

H

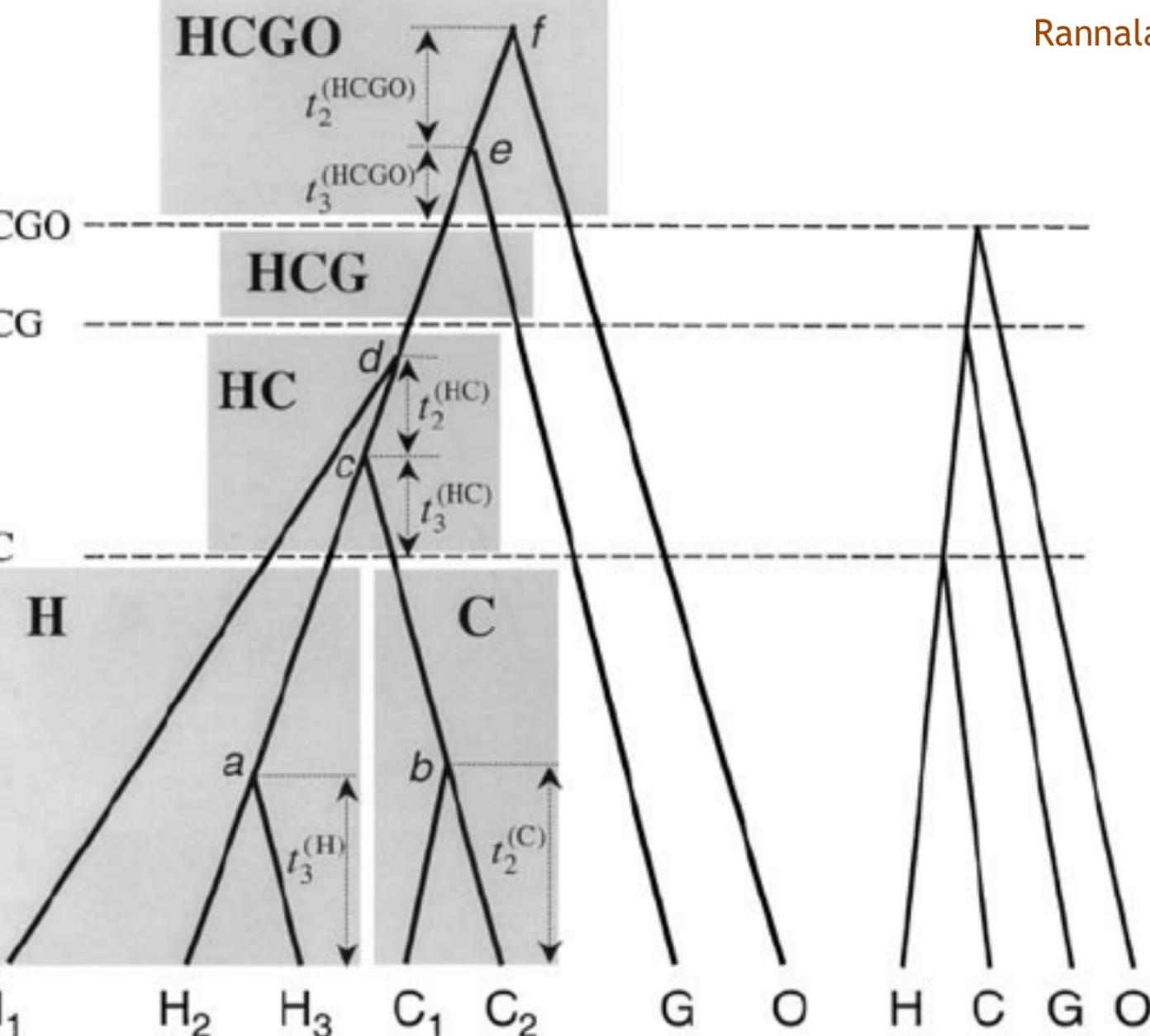
C

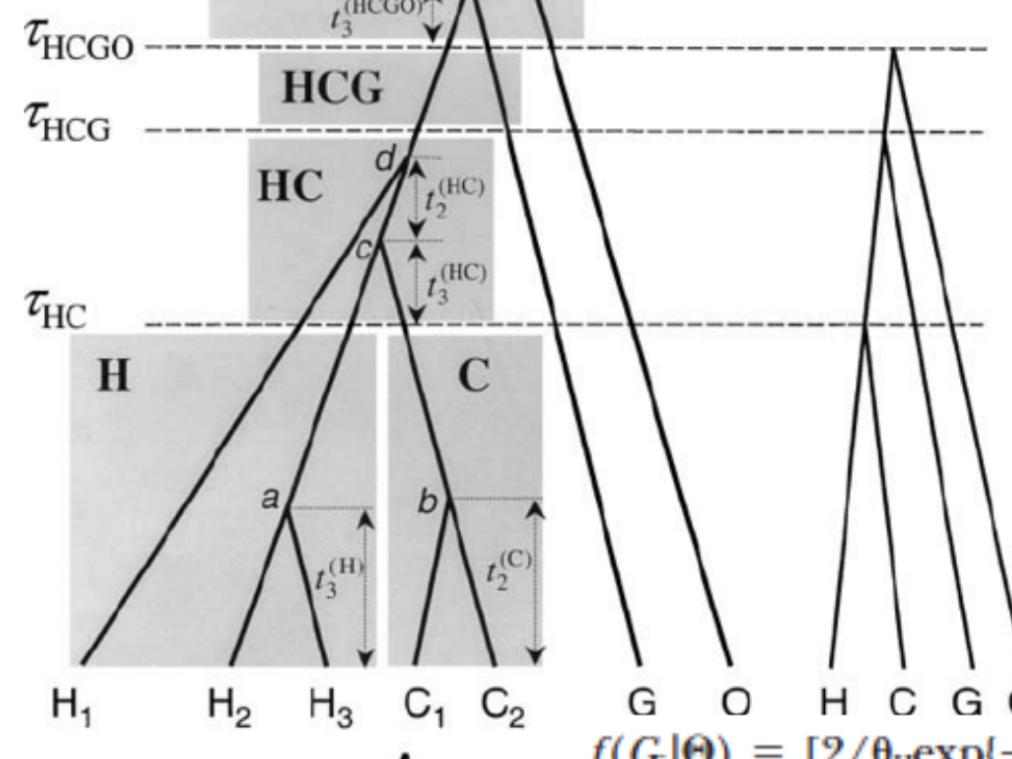
G

O

A

B





Rannala and Yang, *Genetics* 2003

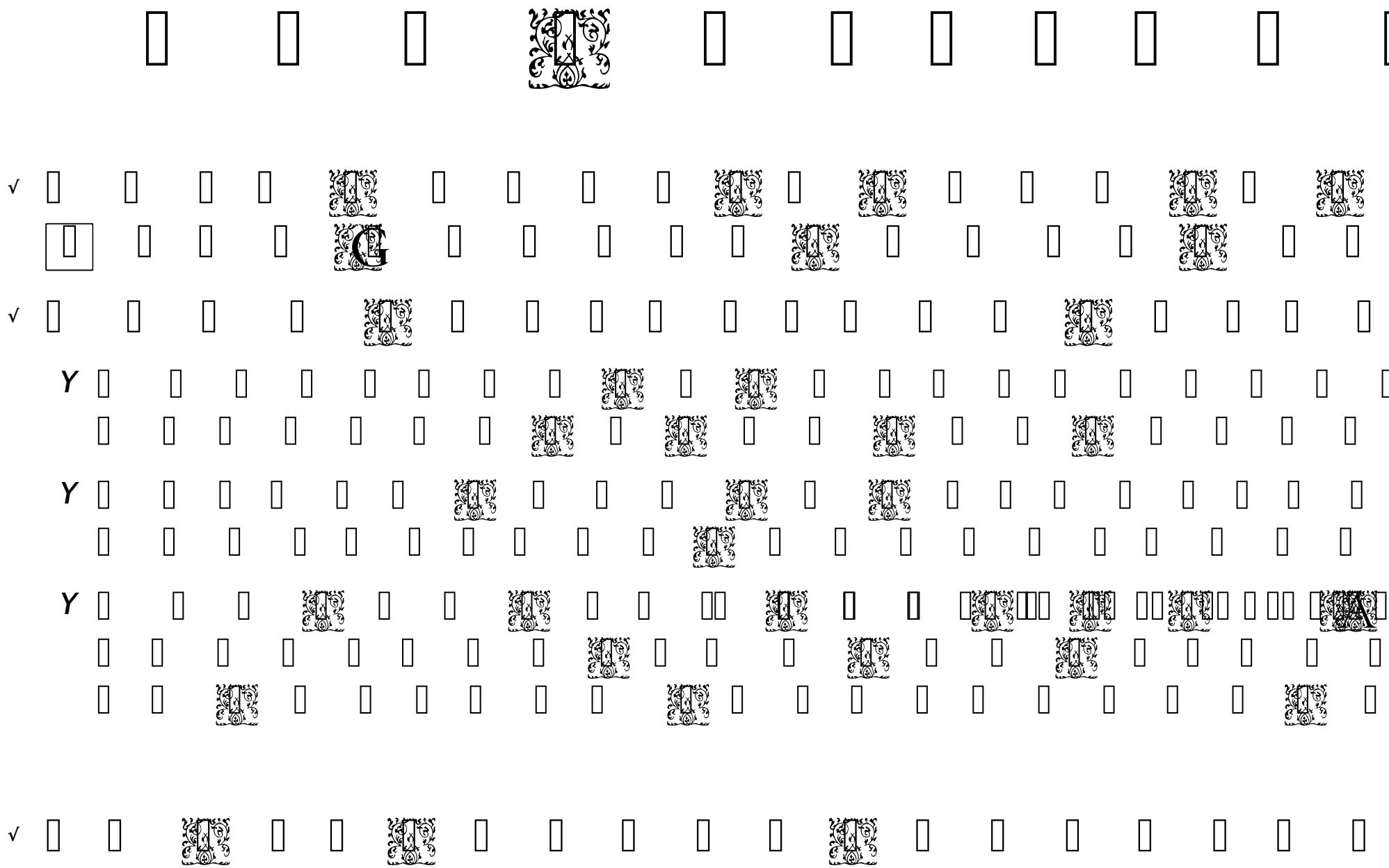
$$f(G_i | \Theta) = [2/\theta_H \exp\{-6t_3^{(H)}/\theta_H\} \exp\{-2(\tau_{HC} - t_3^{(H)})/\theta_H\}]$$

$$\times [2/\theta_C \exp\{-2t_2^{(C)}/\theta_C\}]$$

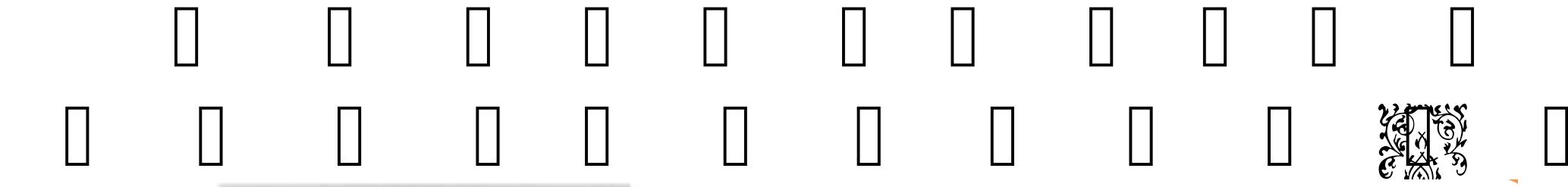
$$\times [2/\theta_{HC} \exp\{-6t_3^{(HC)}/\theta_{HC}\} \times 2/\theta_{HC} \exp\{-2t_2^{(HC)}/\theta_{HC}\}]$$

$$\times [\exp\{-2(\tau_{HCG} - \tau_{HC} - (t_3^{(HC)} + t_2^{(HC)}))/\theta_{HCG}\}]$$

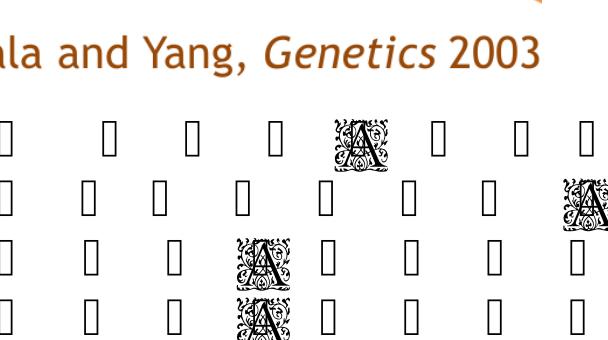
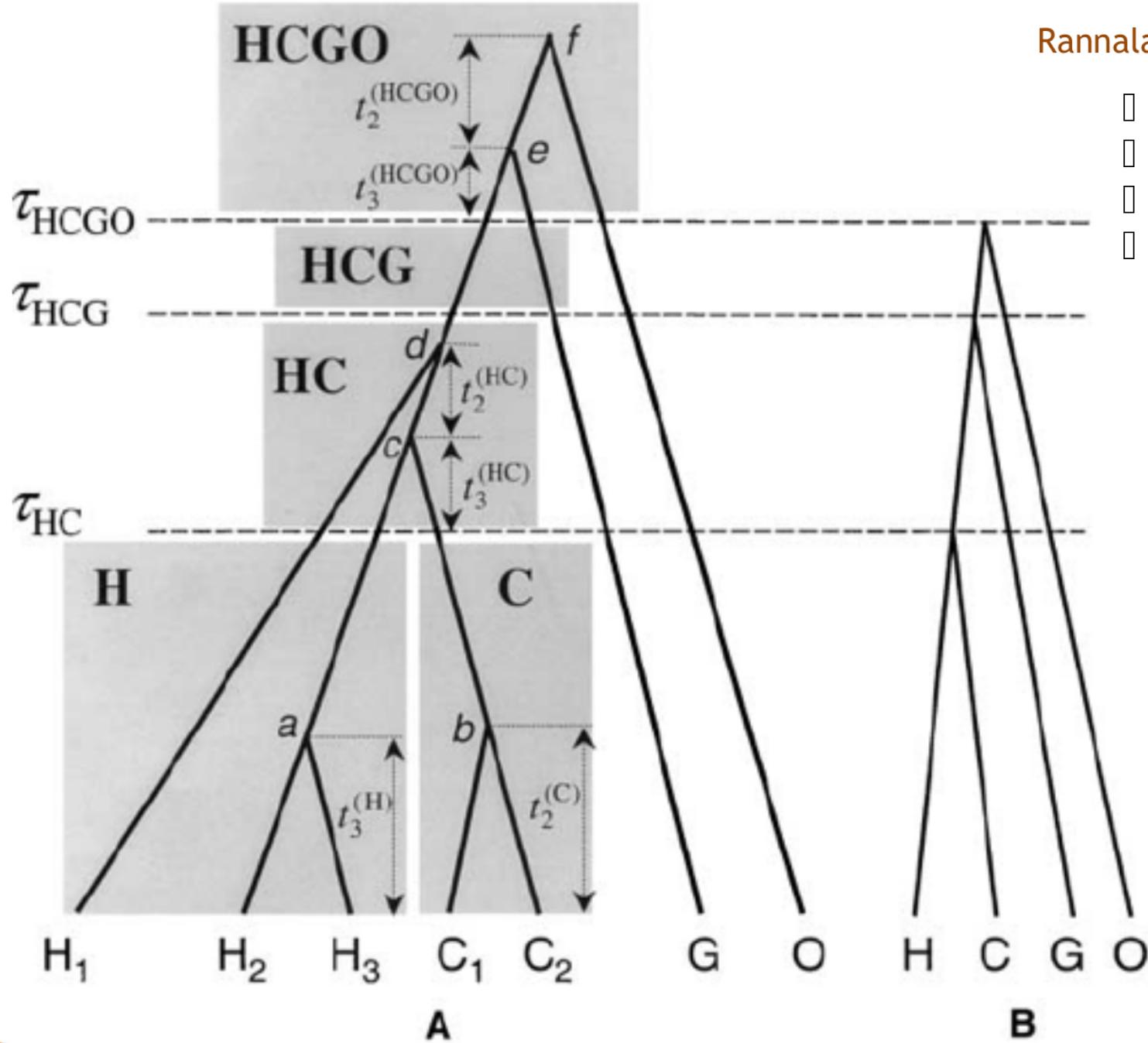
$$\times [2/\theta_{HCGO} \exp\{-6t_3^{(HCGO)}/\theta_{HCGO}\} \times 2/\theta_{HCGO} \exp\{-2t_2^{(HCGO)}/\theta_{HCGO}\}].$$

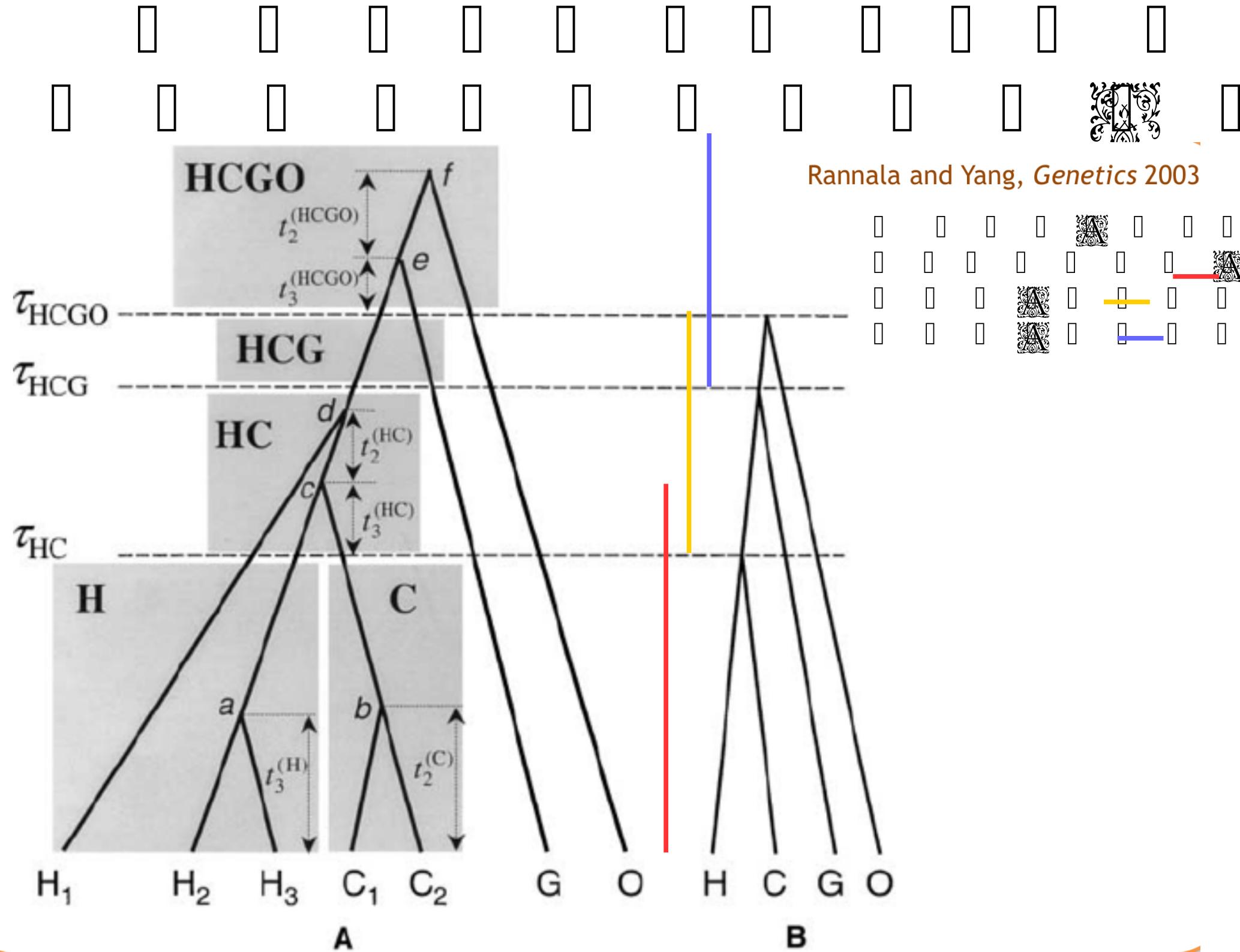


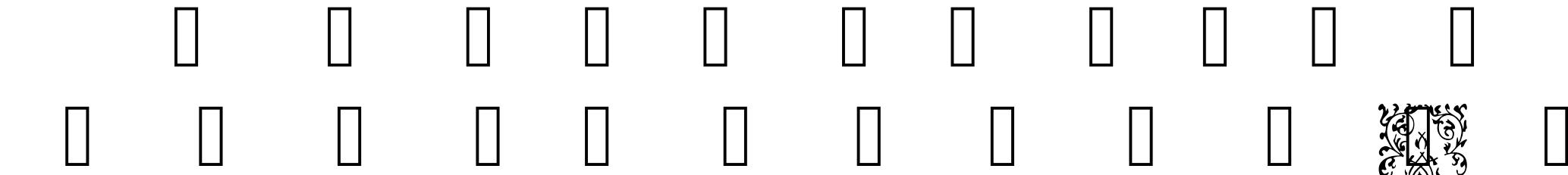
+



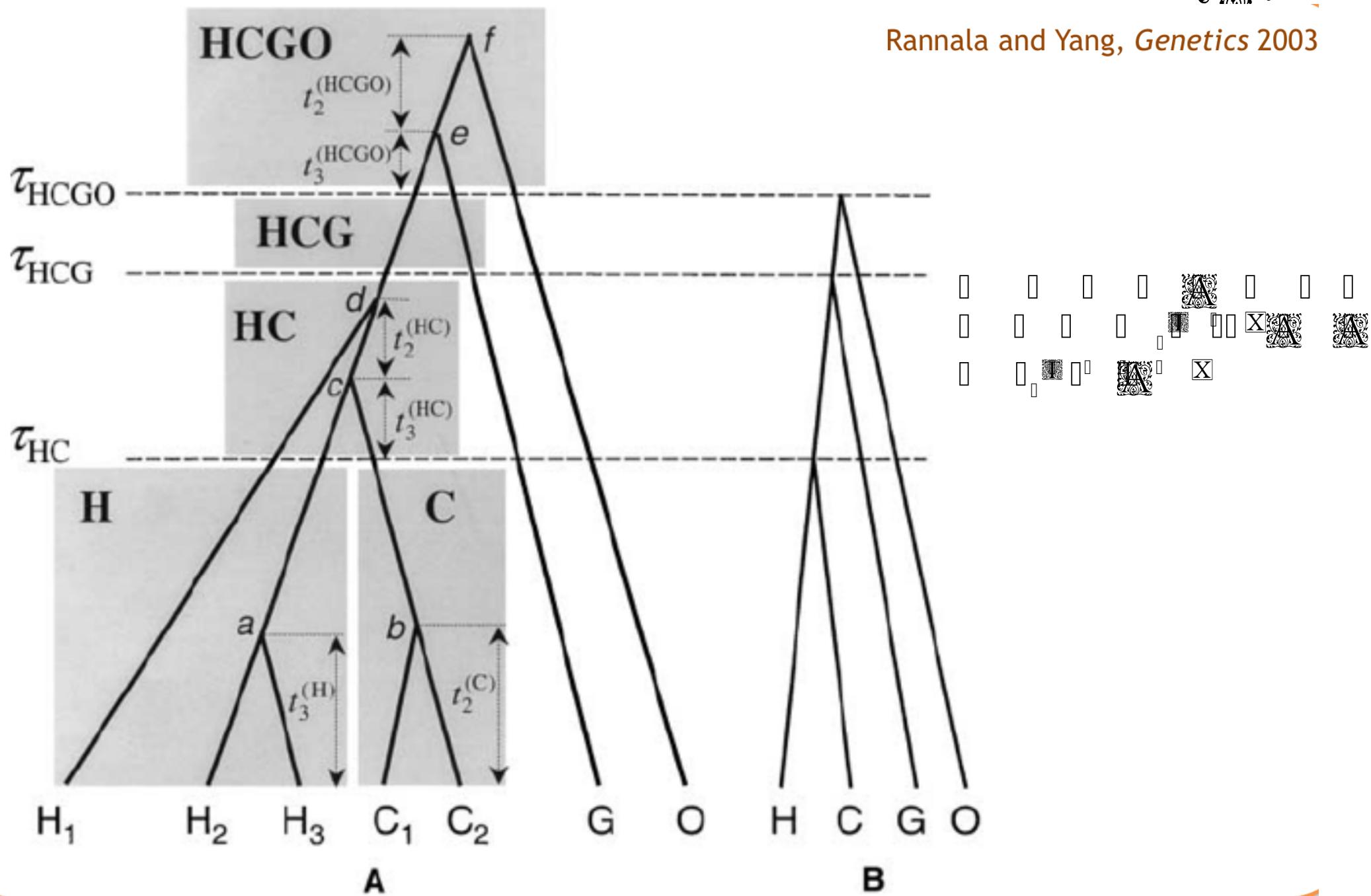
Rannala and Yang, *Genetics* 2003

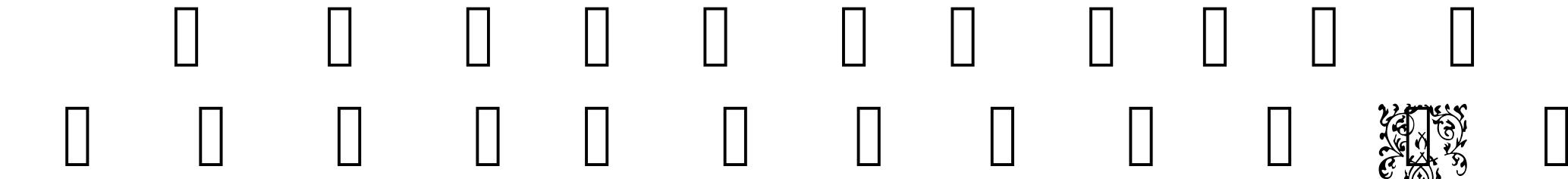




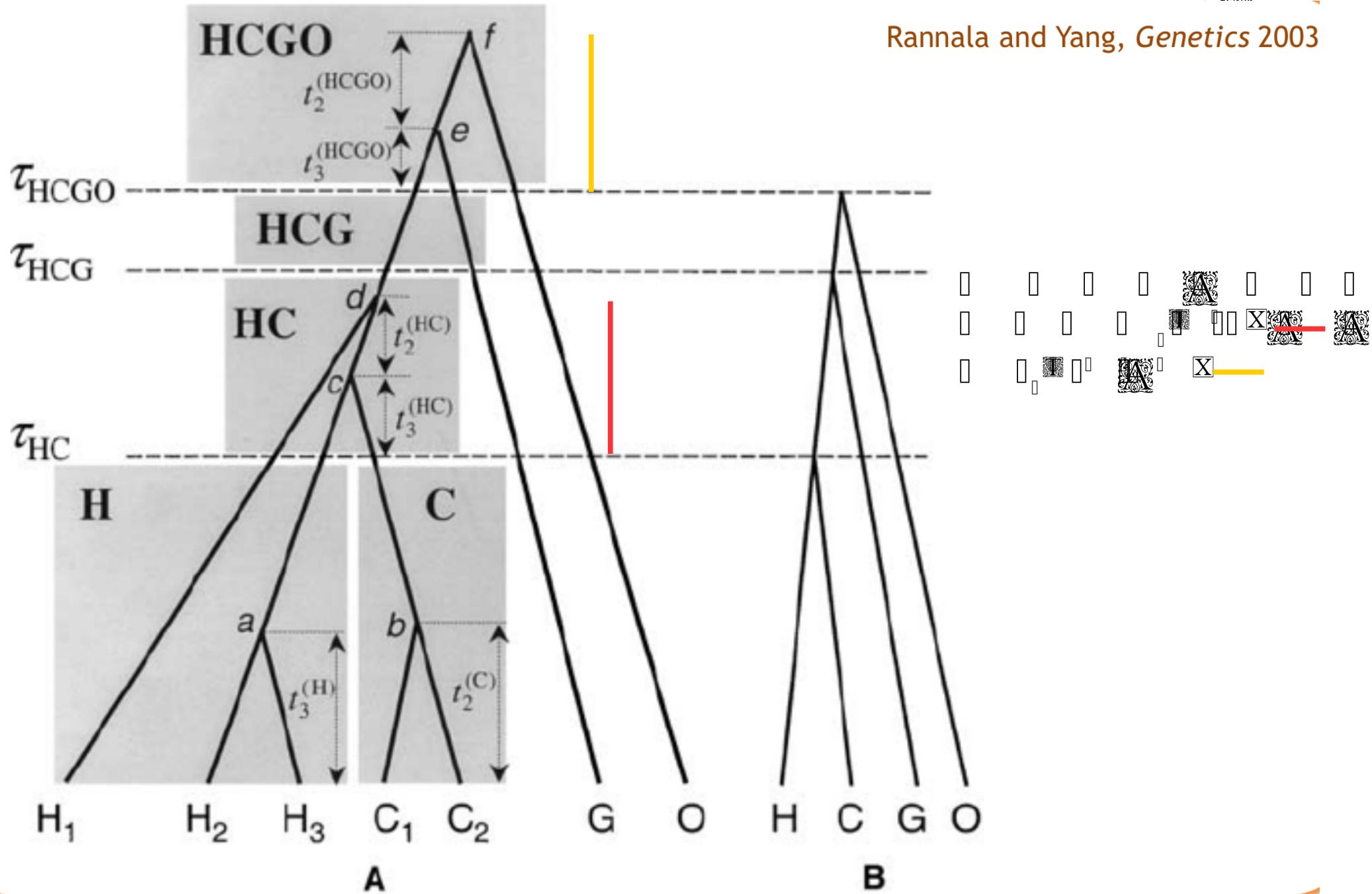


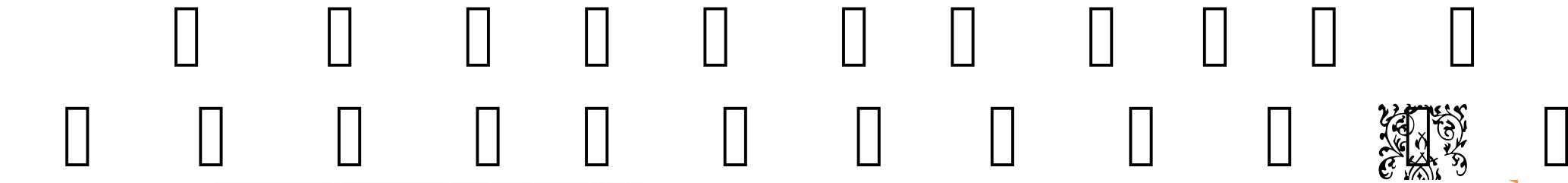
Rannala and Yang, *Genetics* 2003



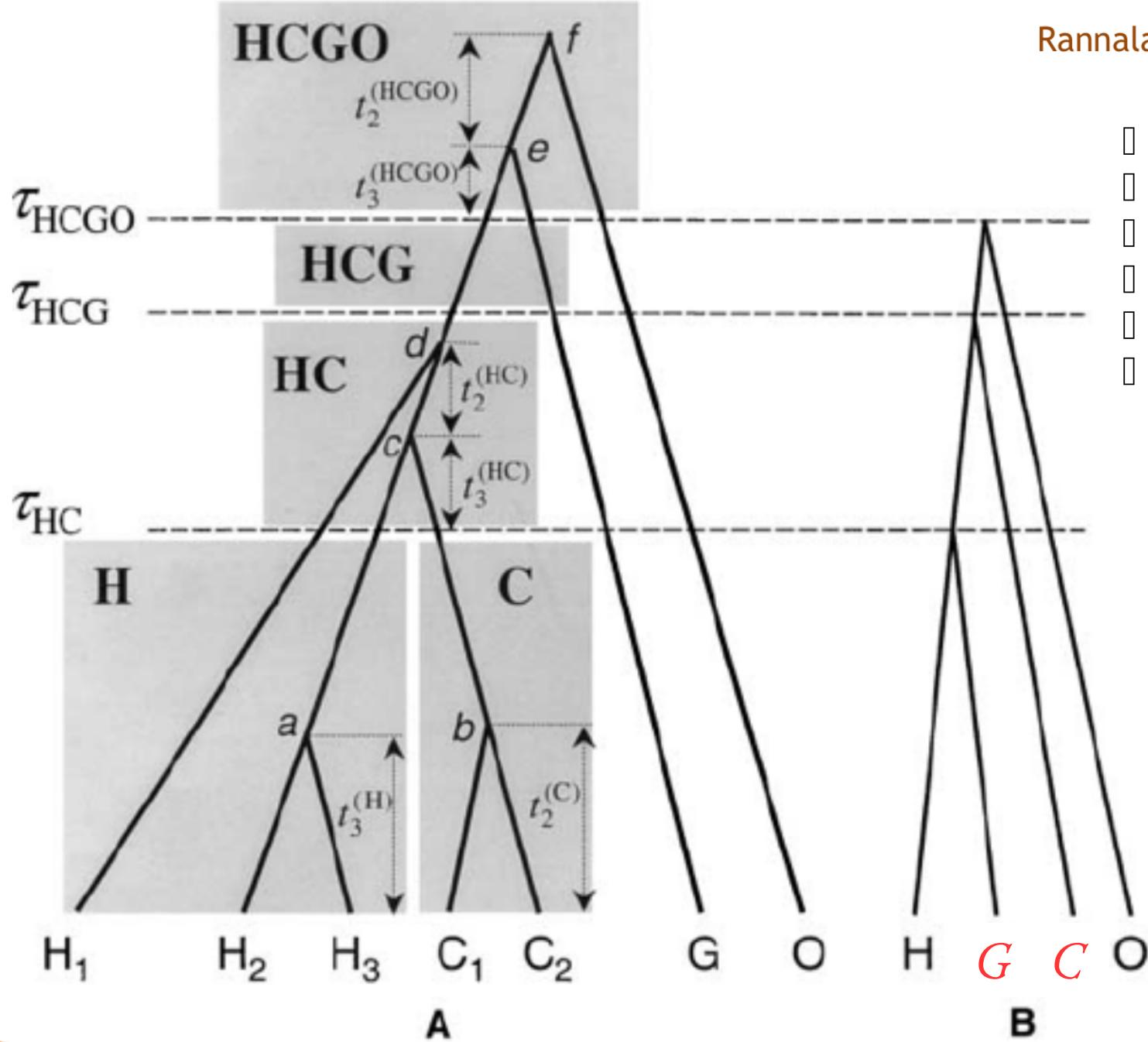


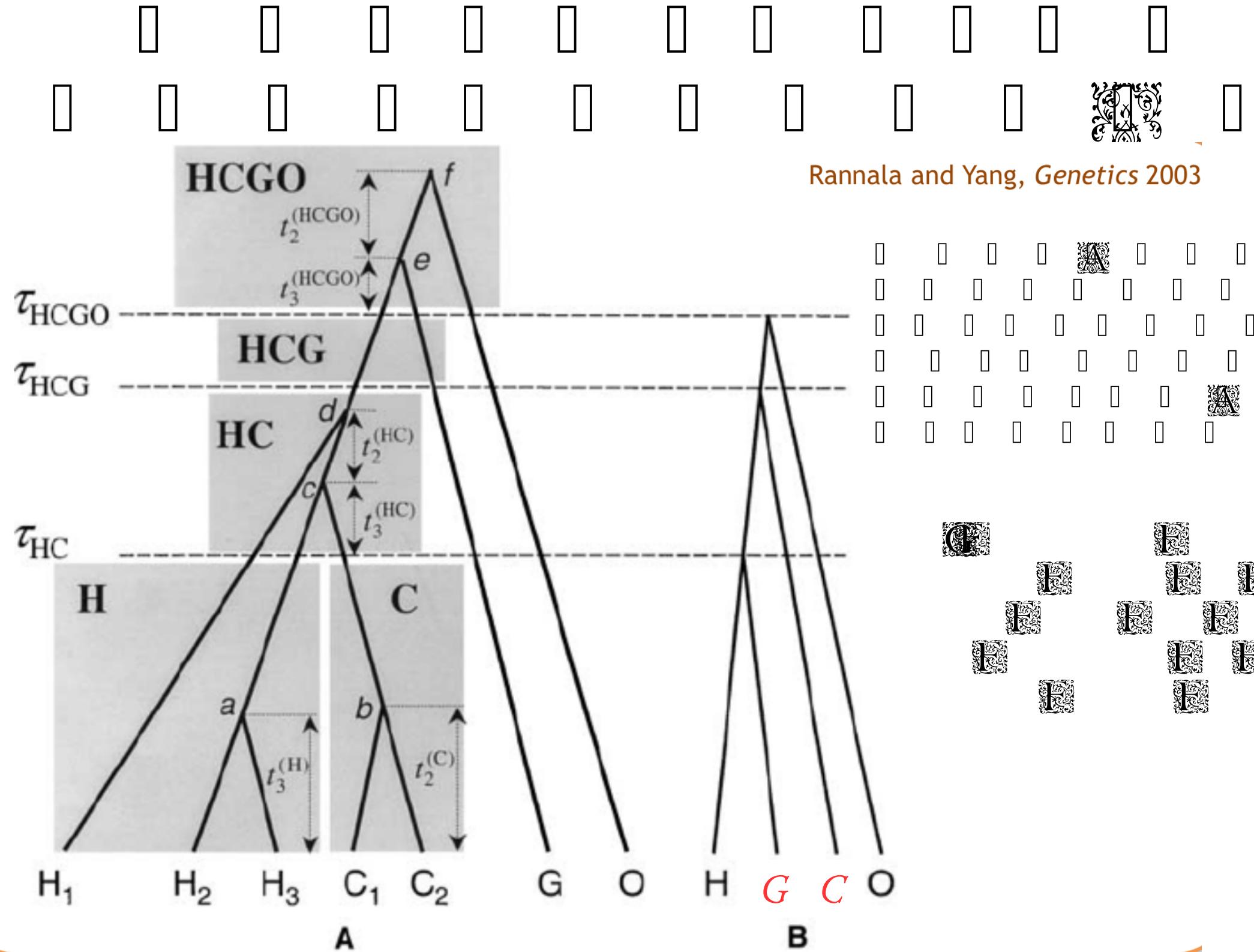
Rannala and Yang, *Genetics* 2003

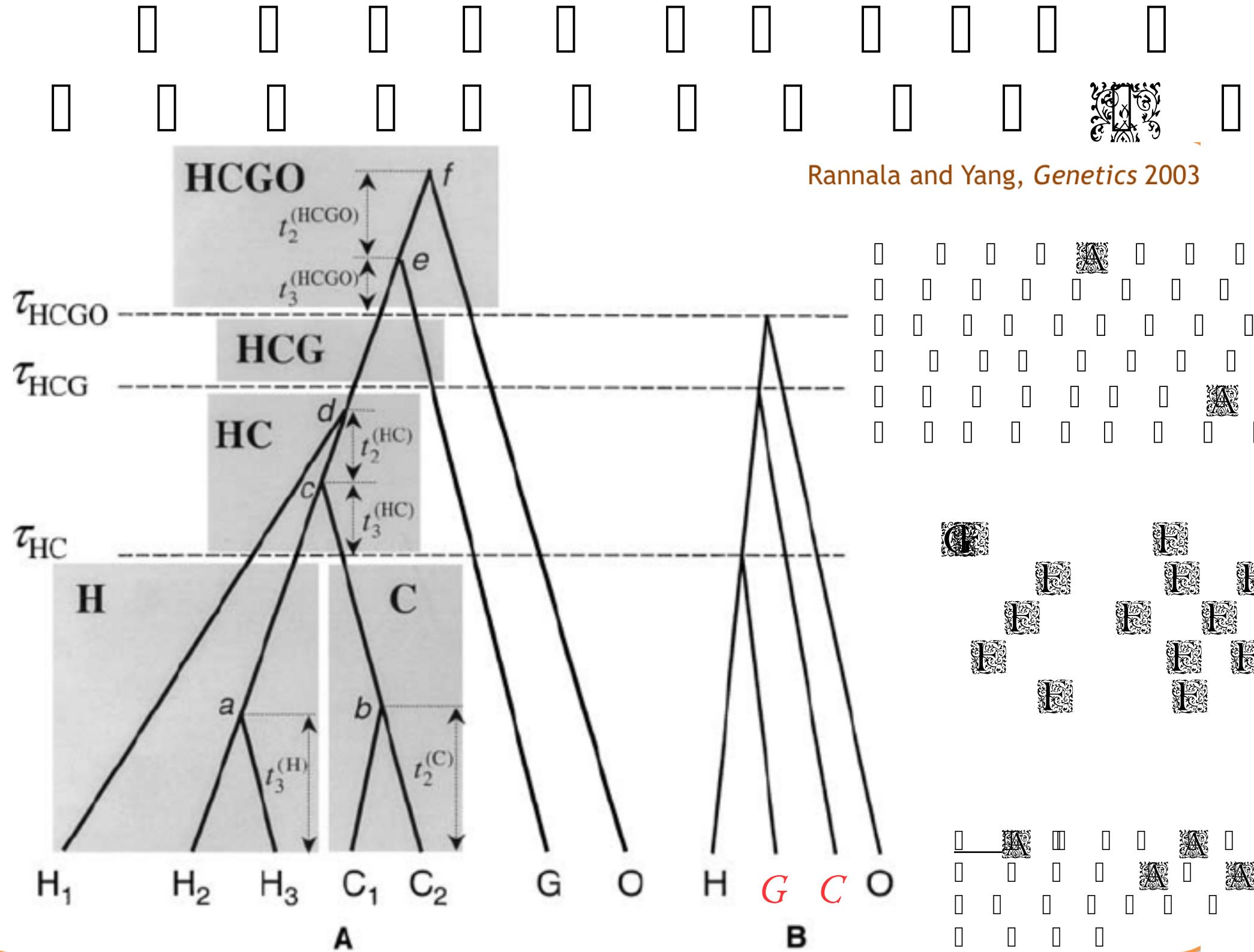


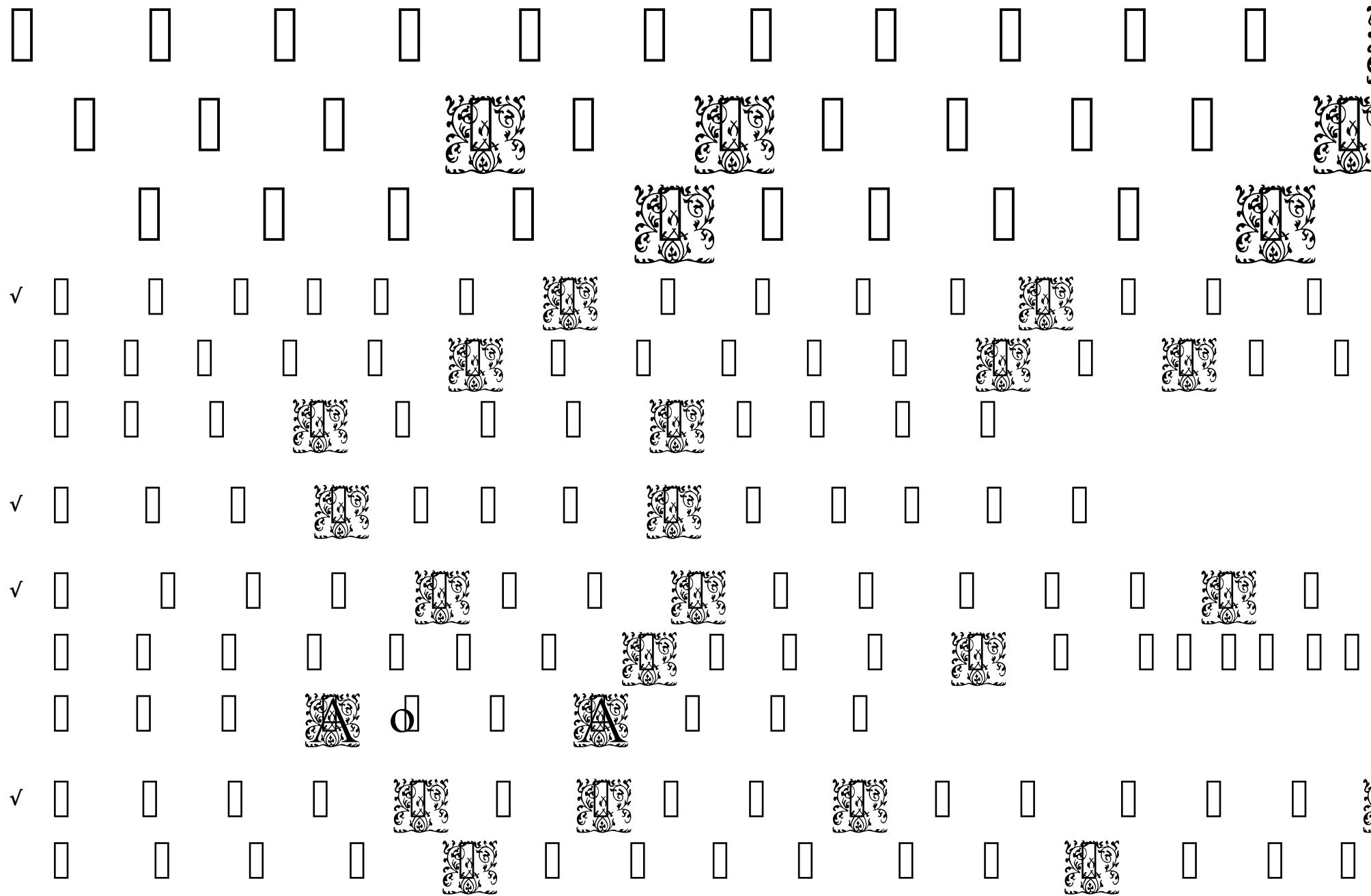


Rannala and Yang, *Genetics* 2003

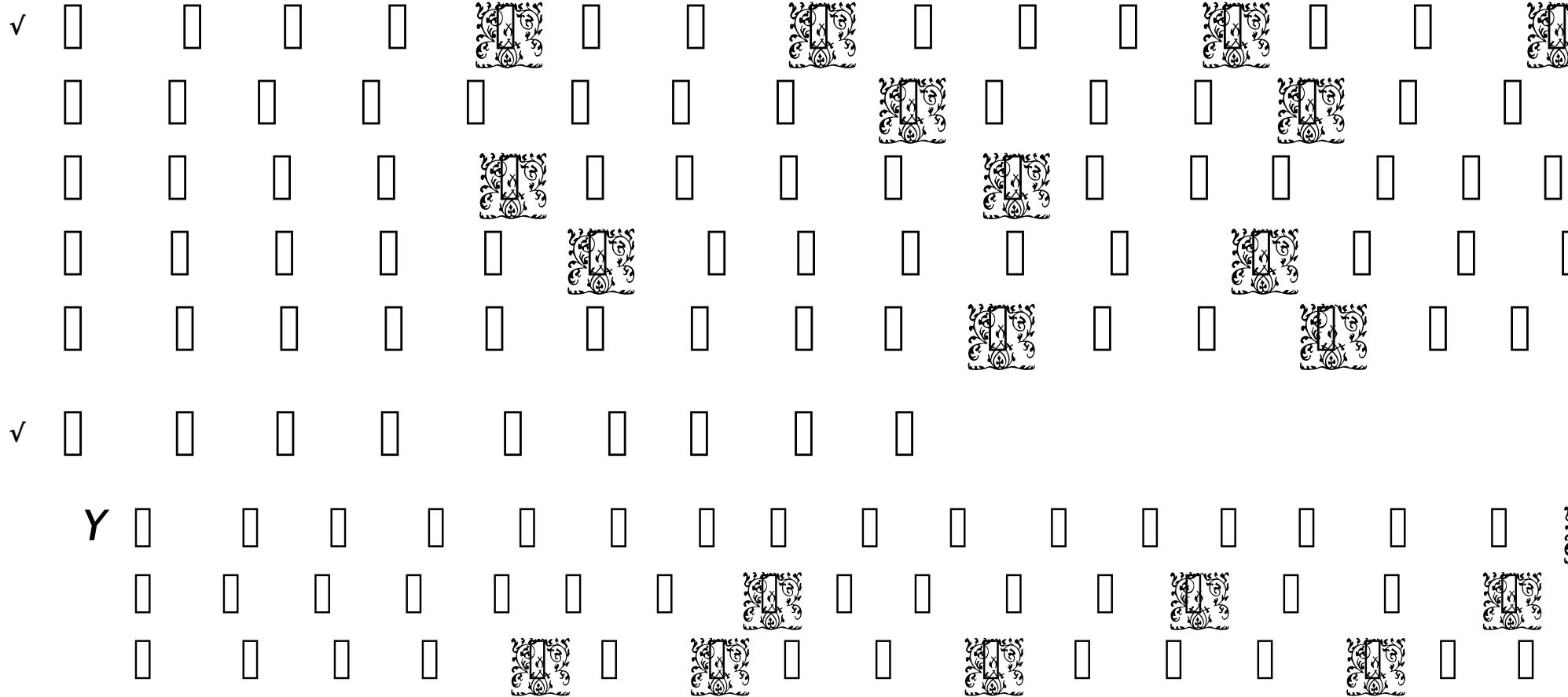
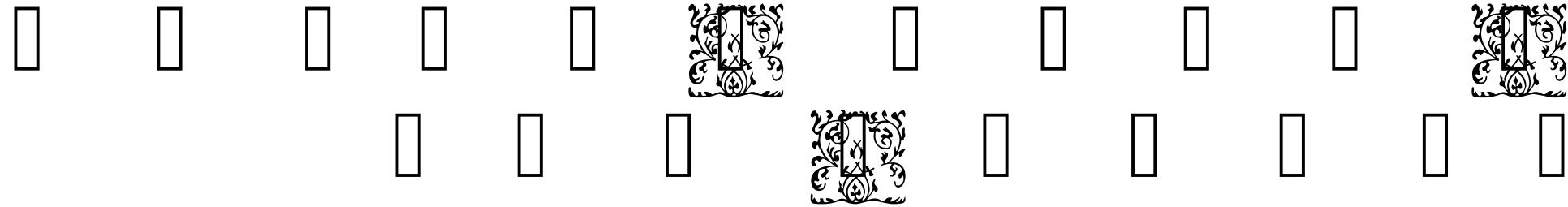








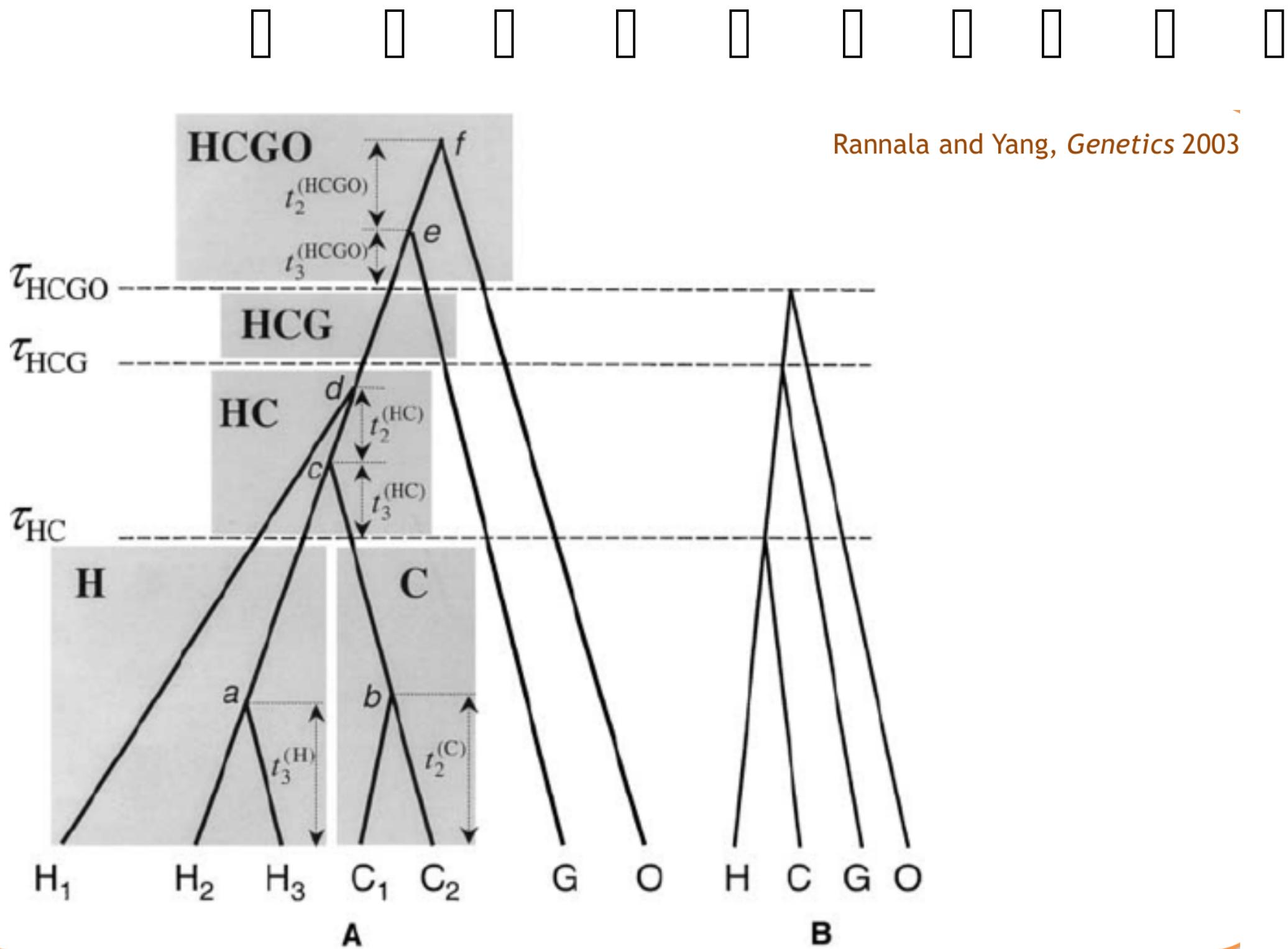
+

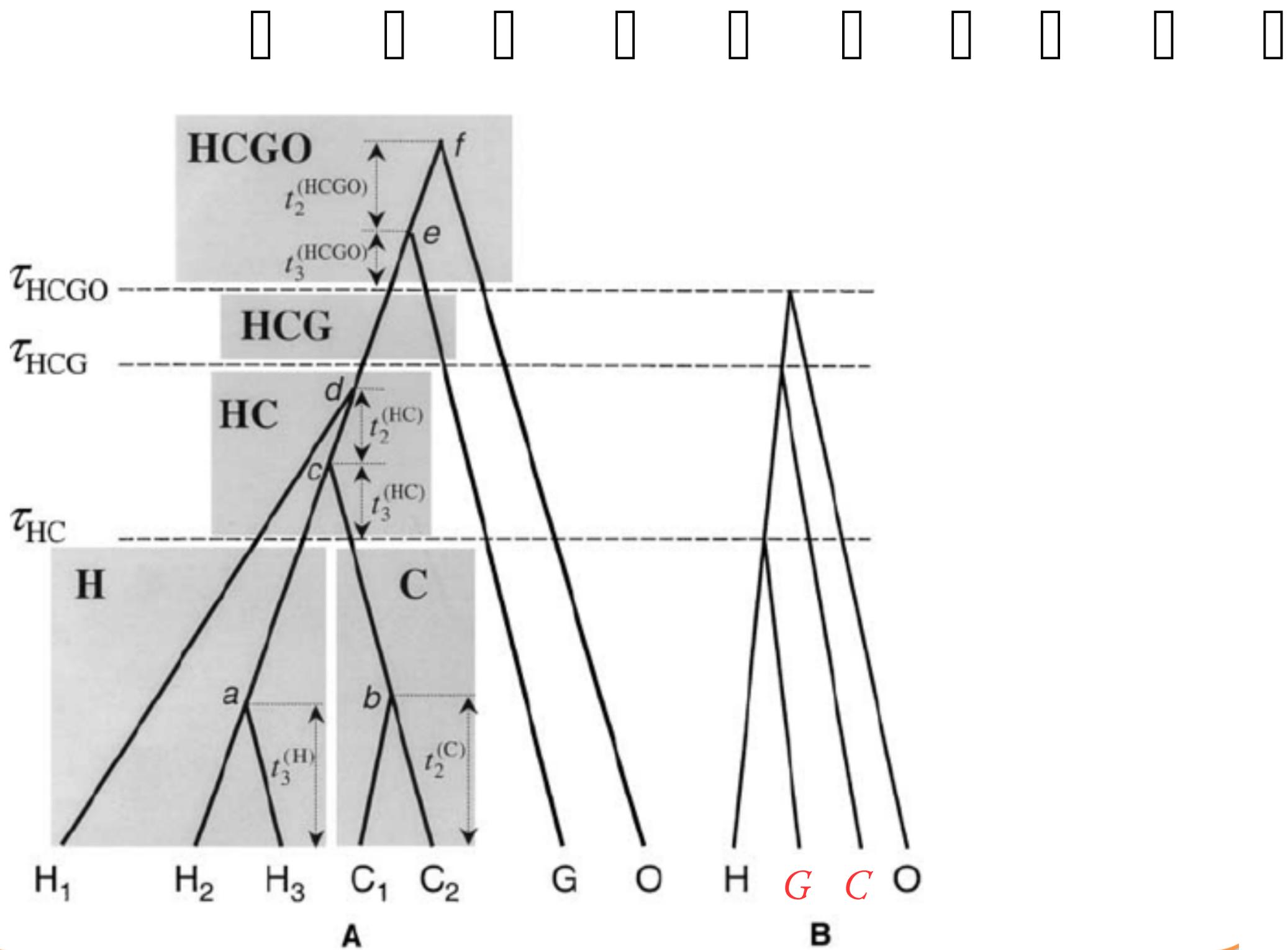


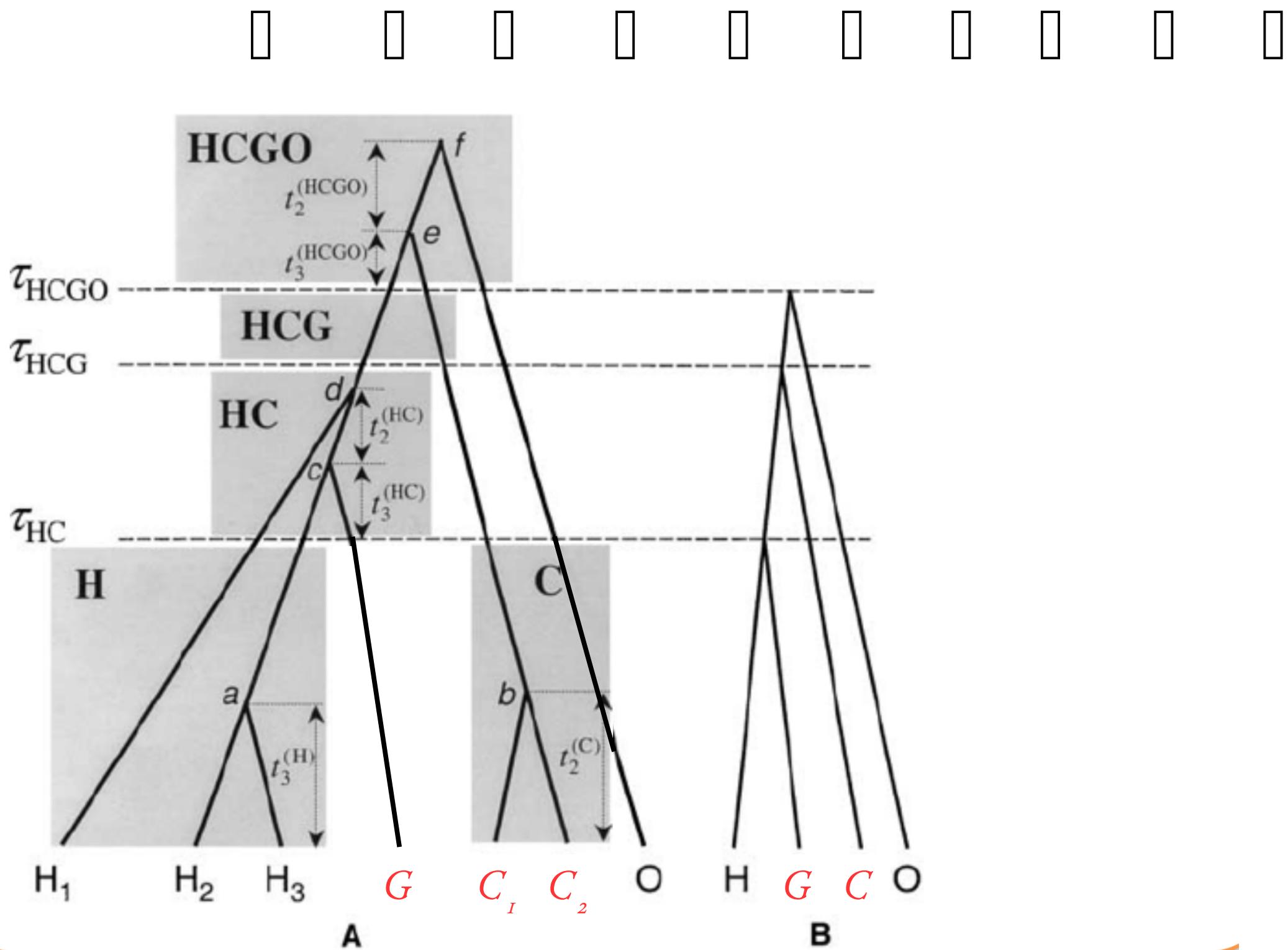
+

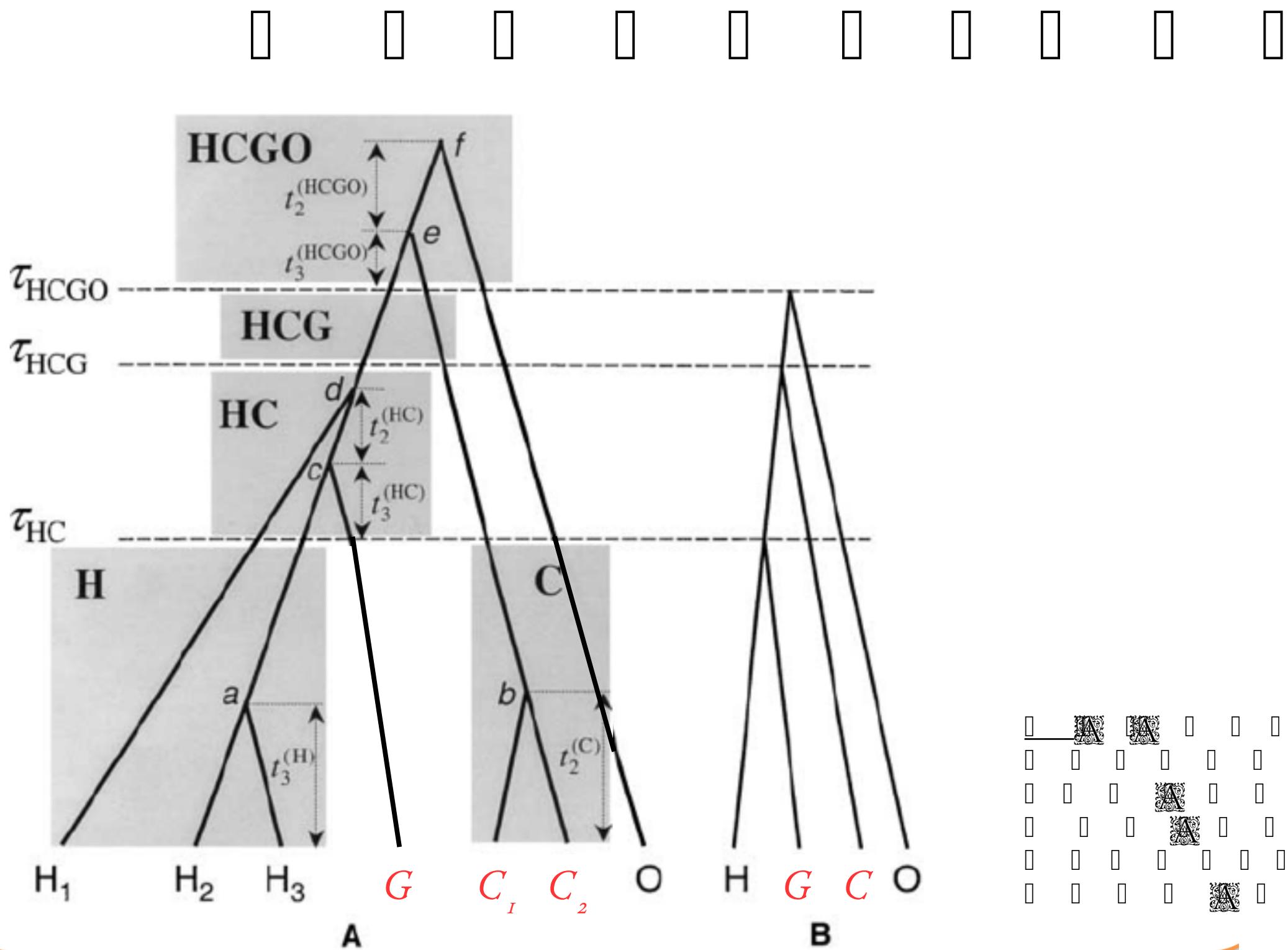
+

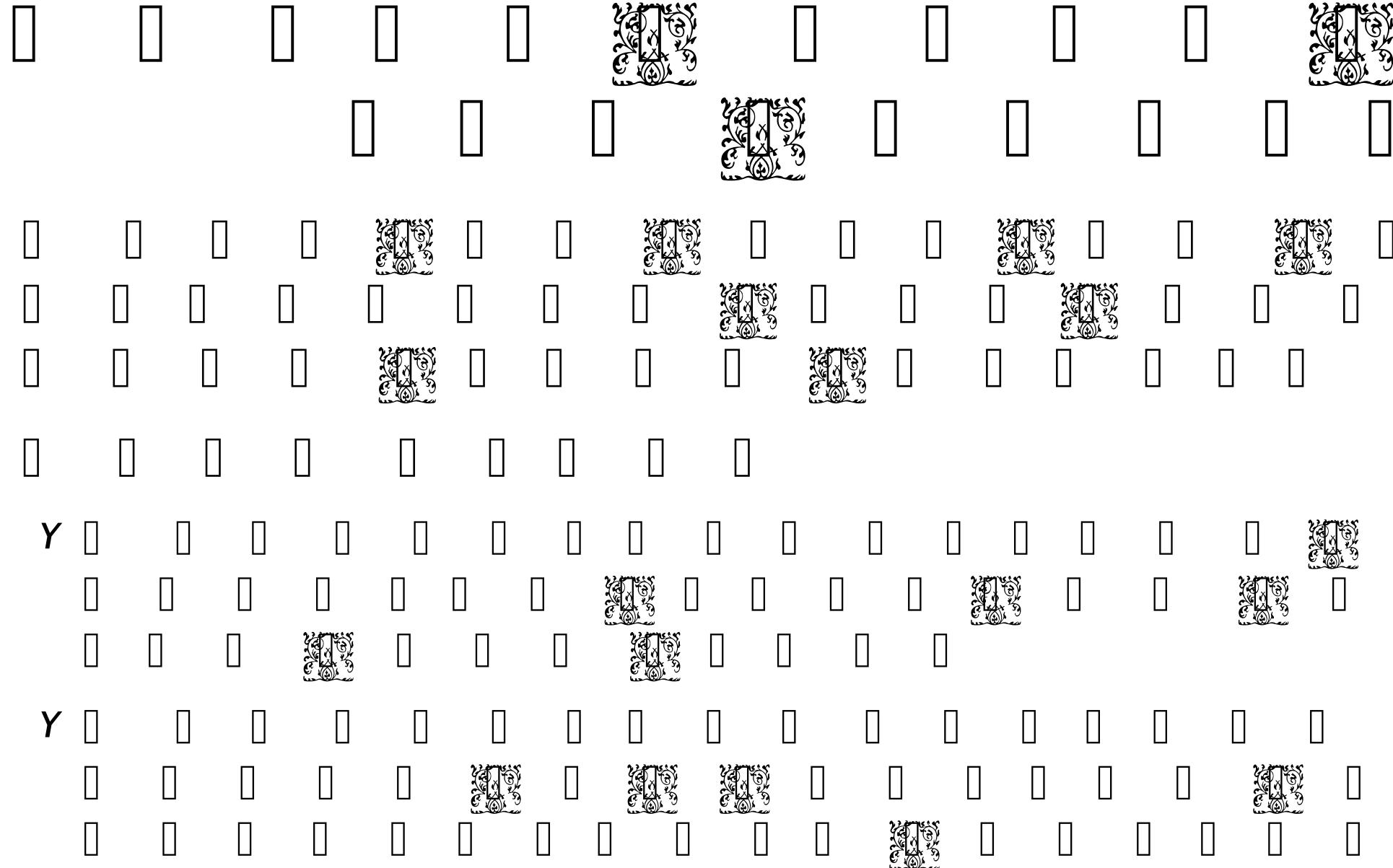
Rannala and Yang, *Genetics* 2003





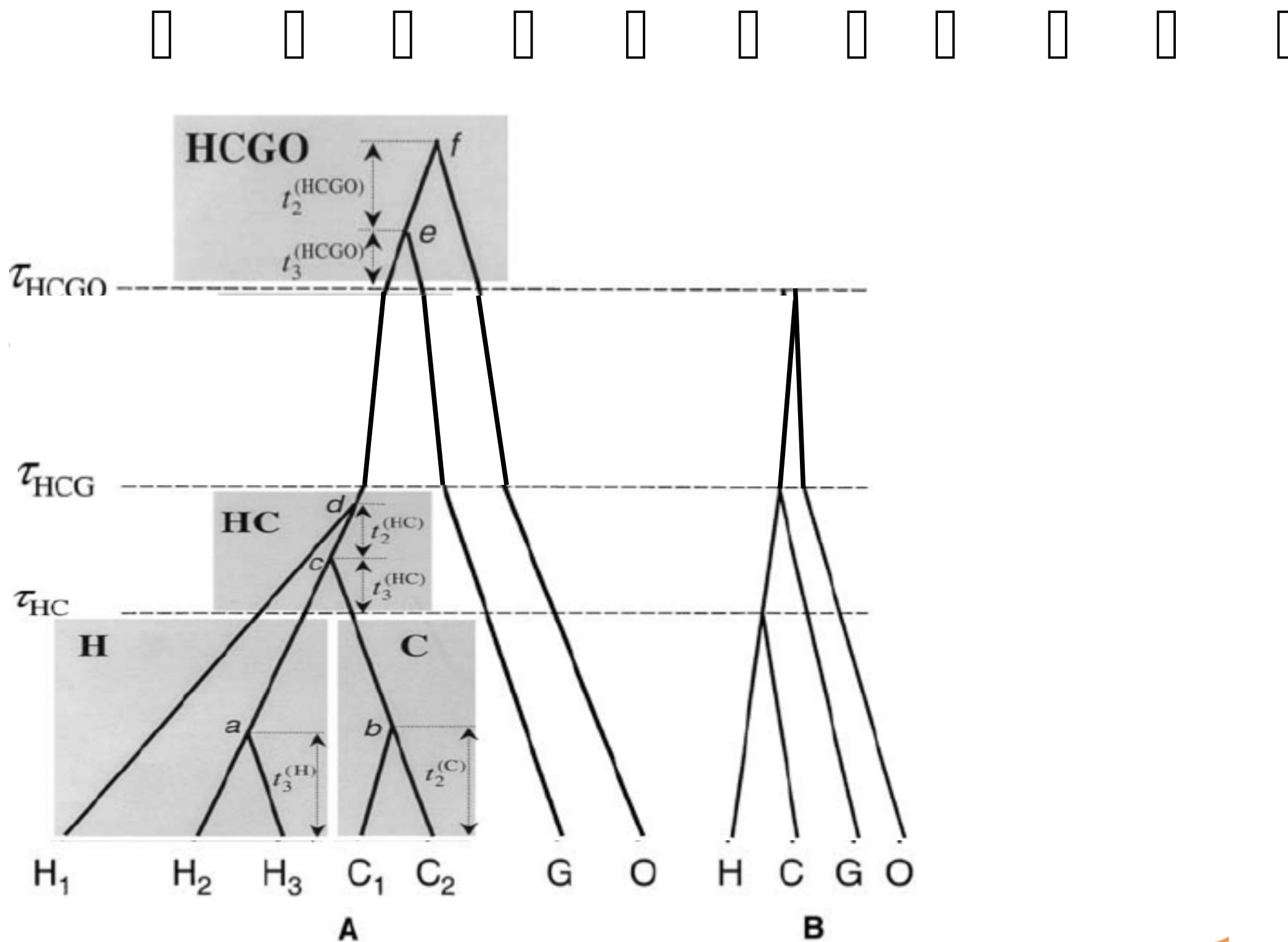


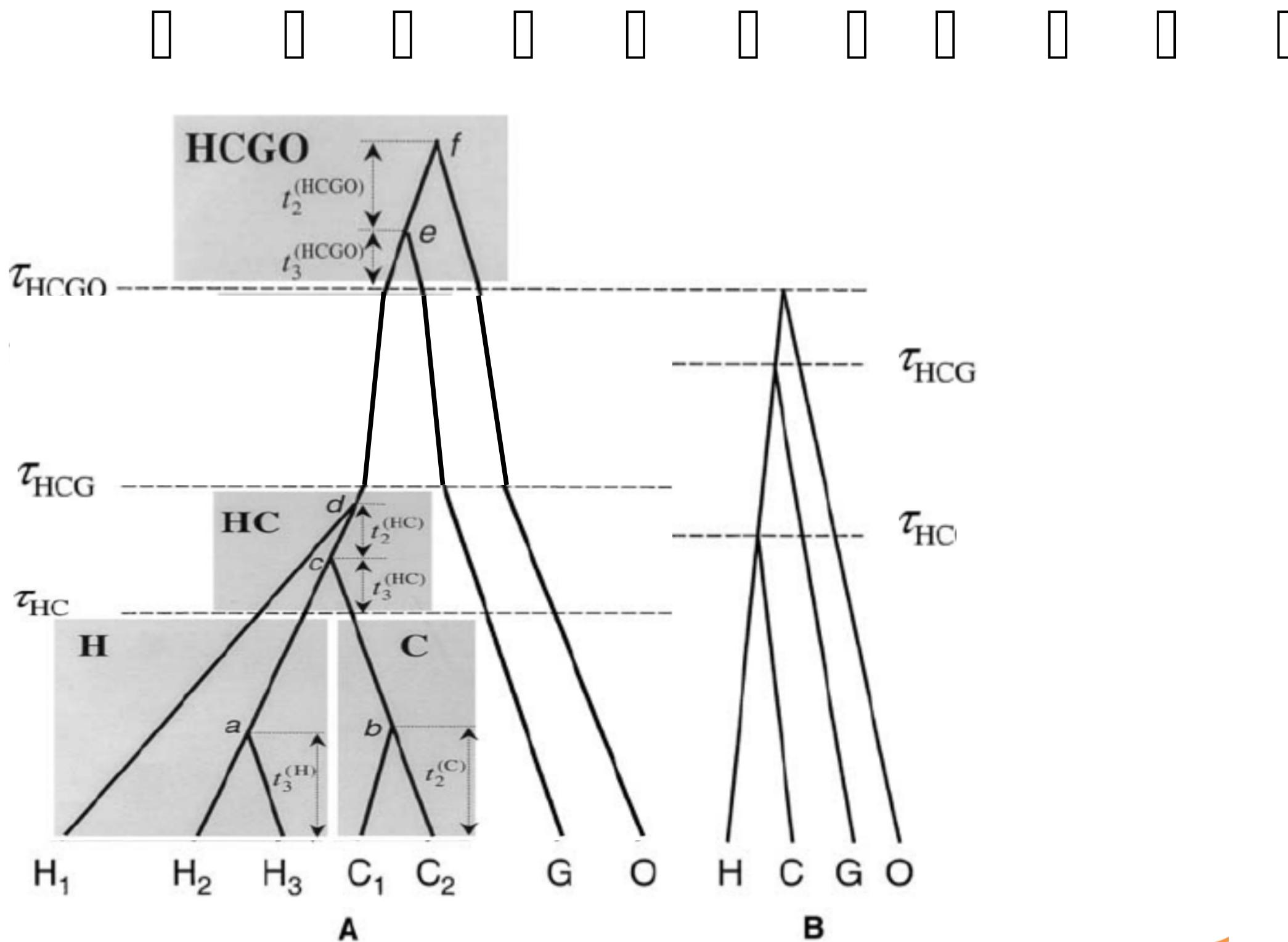


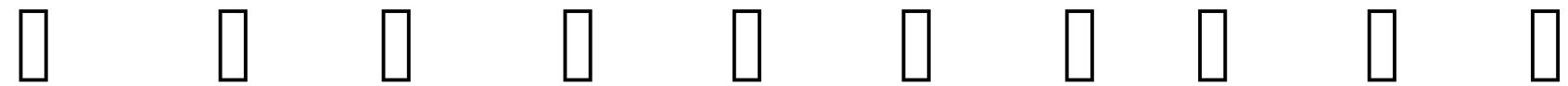


+

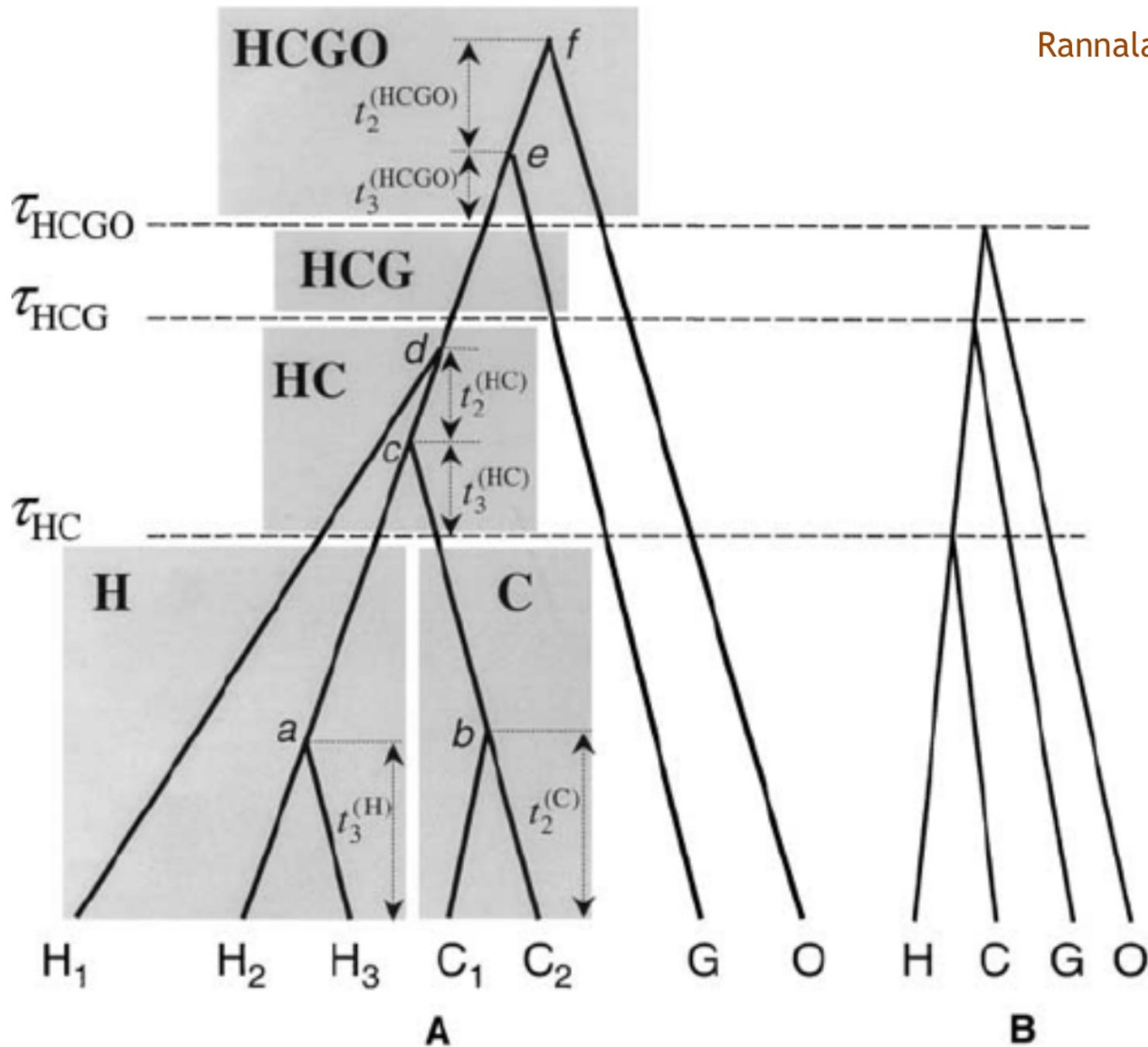
+

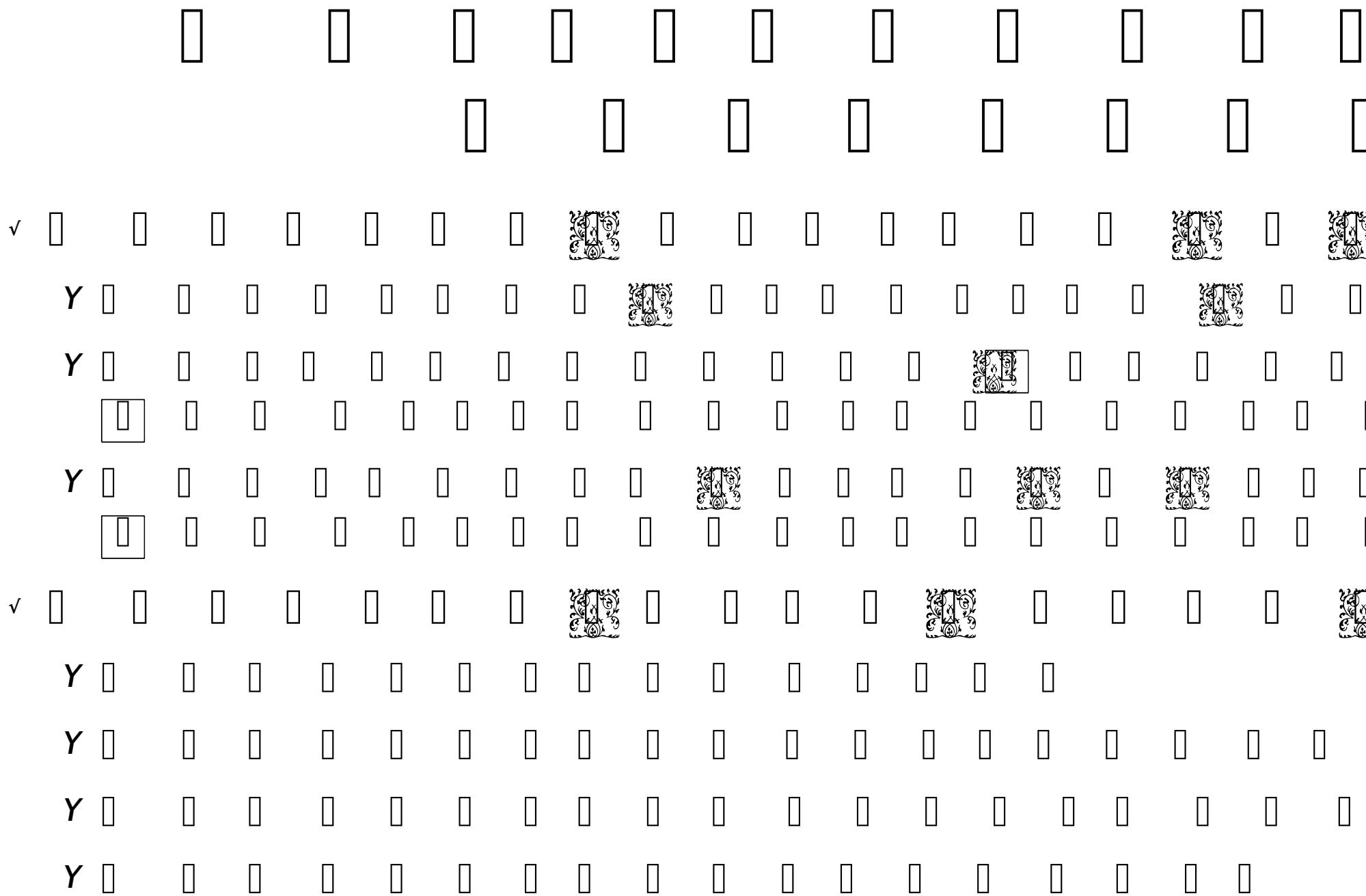






Rannala and Yang, *Genetics* 2003





+

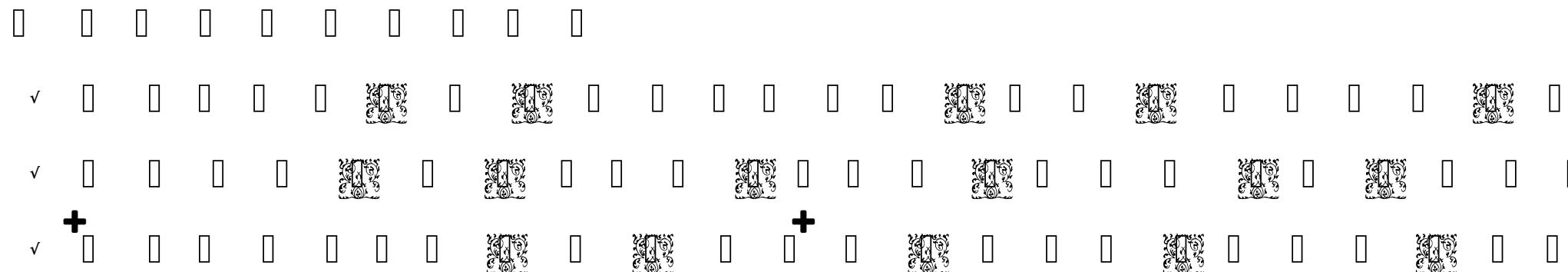
+



```
for ( i in 1 : num_loci ) {  
  taxon_map = readTaxonData( "data/species_maps/primates_" +  
    locus_names[ i ] + "_species_map.txt" )  
  geneTree[ i ] ~ dnMultiSpeciesCoalescent( speciesTree=psi, Ne=Ne,  
    taxa=taxon_map)  
}
```



```
move_species_narrow_exchange = mvSpeciesNarrow( speciesTree=psi ,  
weight=5 )  
  
move_species_subtree_scale = mvSpeciesSubtreeScale( psi , weight=3.0)  
  
for ( i in 1 : numLoci ) {  
  
  move_species_narrow_exchange.addGeneTreeVariable( geneTree[i] )  
  move_species_subtree_scale.addGeneTreeVariable( geneTree[i] )  
  
}  
  
moves[ ++m ] = move_species_narrow_exchange  
moves[ ++m ] = move_species_subtree_scale
```



+

v

v

Y

Y

Y

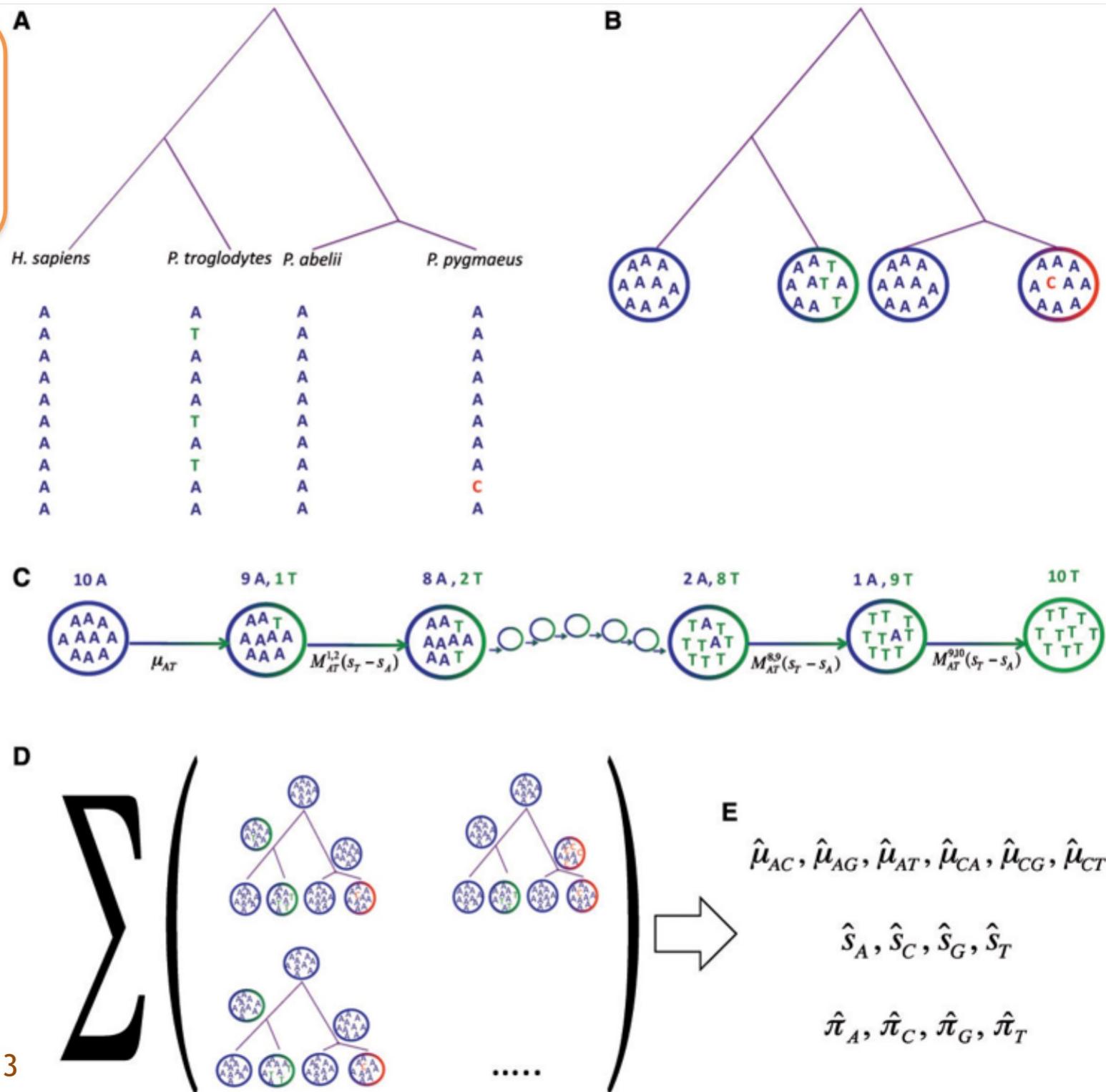


+

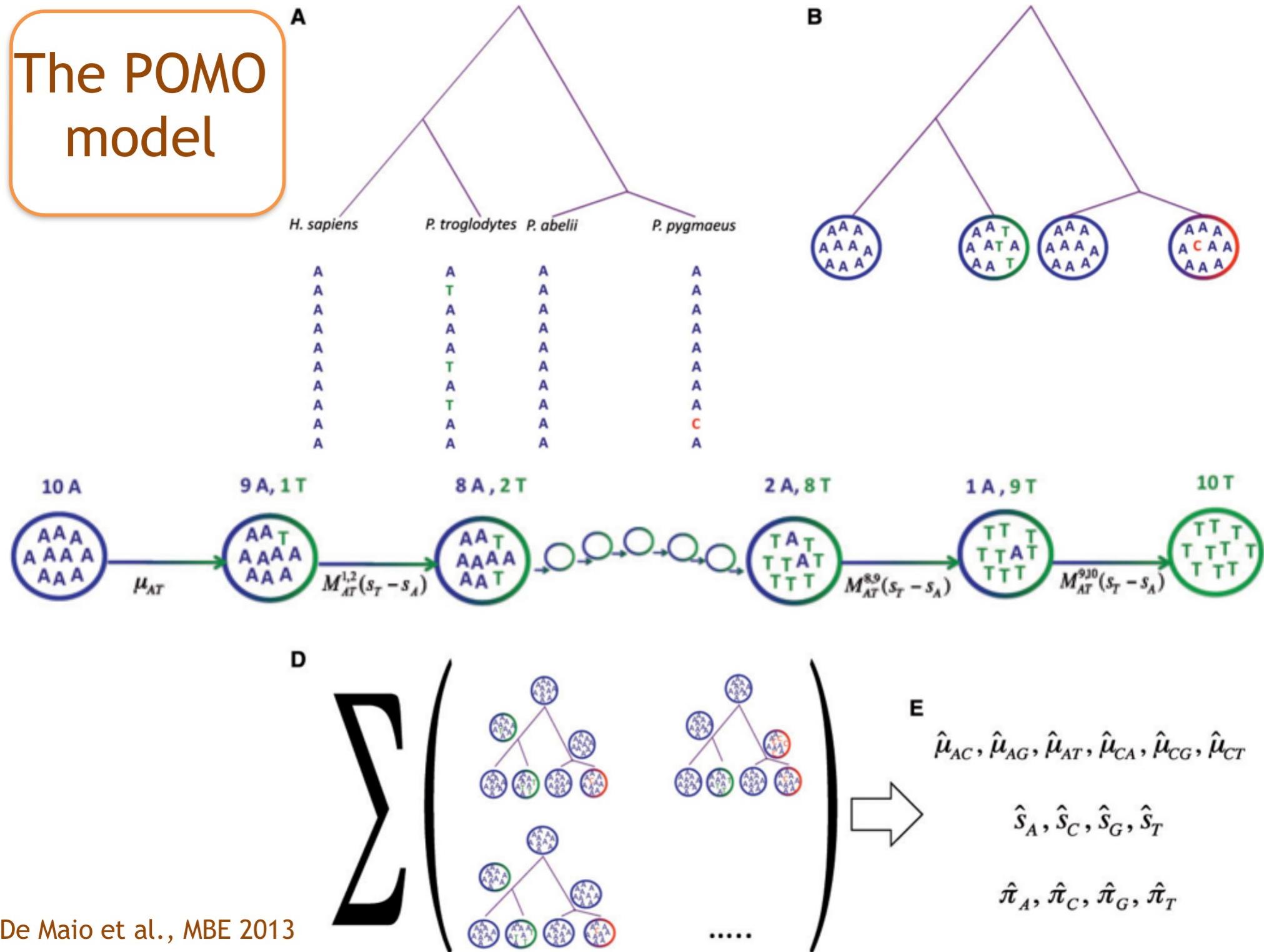
Reconstructing species trees given ILS

- Parsimony method: minimizing deep coalescences (Maddison 1997)
- Distance method: NJst (Liu and Yu 2011)
- Summary statistic methods: STAR (Liu et al, 2009), GLASS (Mossel and Roch 2010), iGLASS (Jewett and Rosenberg 2012), ASTRAL (Mirarab et al., 2014)
- Maximum Likelihood methods: STEM (Kubatko et al. 2009), MP-EST (Liu et al., 2010)
- Bayesian methods: BeST (2007), *BEAST (Heled and Drummond, 2009)
- Bypassing the gene trees with math tricks: SNAPP (Bryant et al., 2012), POMO (DeMaio et al., 2013), SVDQuartets (Chifman and Kubatko, 2014)

The POMO model



The POMO model



The POMO model

$$\begin{array}{l}
 \mathbf{A} \\
 \left(\begin{array}{c} \binom{1}{N-1} \\ \binom{1}{N-2} \\ \binom{2}{N-2} \\ \binom{3}{N-3} \\ \binom{4}{N-4} \\ \vdots \\ \binom{N-4}{N-4} \\ \binom{N-3}{N-3} \\ \binom{N-2}{N-2} \\ \binom{N-1}{N-1} \end{array} \right) \left(\begin{array}{ccccccccc} * & \binom{2}{N-2} & \binom{3}{N-3} & \binom{4}{N-4} & \cdots & \binom{N-4}{4} & \binom{N-3}{3} & \binom{N-2}{2} & \binom{N-1}{1} \\ N \cdot M_{IJ}^{1,2} & * & N \cdot M_{IJ}^{2,3} & & & & & & \\ N \cdot M_{IJ}^{2,1} & N \cdot M_{IJ}^{3,2} & * & N \cdot M_{IJ}^{3,4} & & & & & 0 \\ N \cdot M_{IJ}^{3,1} & & & & & & & & \\ N \cdot M_{IJ}^{4,3} & & & * & & & & & \\ \ddots & & & \ddots & & & & & \\ \ddots & & & \ddots & * & N \cdot M_{IJ}^{N-3,N-2} & N \cdot M_{IJ}^{N-2,N-3} & N \cdot M_{IJ}^{N-2,N-1} & N \cdot M_{IJ}^{N-1,N} \\ 0 & & & & N \cdot M_{IJ}^{N-2,N-3} & * & N \cdot M_{IJ}^{N-1,N-2} & * & N \cdot M_{IJ}^{N-1,N} \\ 0 & & & & & N \cdot M_{IJ}^{N-1,N-2} & * & N \cdot M_{IJ}^{N-1,N} & * \\ 0 & & & & & & N \cdot M_{IJ}^{N,N-1} & & N \cdot M_{IJ}^{N,N-1} \\ 0 & & & & & & & & * \end{array} \right) \\
 \mathbf{B} \\
 \left(\begin{array}{c} A \\ C \\ G \\ T \\ \binom{1}{N-1} \\ \vdots \\ \binom{N-1}{N-1} \\ \binom{1}{N-1} \\ \binom{1}{N-1} \\ \binom{1}{N-1} \\ \vdots \\ \binom{1}{N-1} \\ \vdots \\ \vdots \end{array} \right) \left(\begin{array}{cccc|cccc|cccc|cccc|} A & C & G & T & \binom{1}{N-1} & \cdots & \binom{N-1}{1} & \binom{1}{N-1} & \cdots & \binom{N-1}{1} & \binom{1}{N-1} & \cdots & \binom{N-1}{1} & \cdots & \cdots \\ * & 0 & 0 & 0 & 0 & \cdots & N^2 \cdot \mu_{AC} & 0 & \cdots & N^2 \cdot \mu_{AG} & 0 & \cdots & N^2 \cdot \mu_{AT} & \cdots & N^2 \cdot \mu_{AT} \\ 0 & * & 0 & 0 & 0 & \cdots & 0 & 0 & \cdots & 0 & 0 & \cdots & 0 & \cdots & 0 \\ 0 & 0 & * & 0 & 0 & \cdots & 0 & 0 & \cdots & 0 & 0 & \cdots & 0 & \cdots & 0 \\ 0 & 0 & 0 & * & 0 & \cdots & 0 & 0 & \cdots & 0 & 0 & \cdots & 0 & \cdots & 0 \\ \hline N \cdot M_{AC}^{1,0} & 0 & 0 & 0 & 0 & \cdots & B_{AC}^N & 0 & \cdots & 0 & 0 & \cdots & 0 & \cdots & 0 \\ \vdots & & & & & \cdots & & & \cdots & & & \cdots & & & \cdots \\ 0 & N \cdot M_{AC}^{N,N+1} & 0 & 0 & 0 & \cdots & B_{AC}^N & 0 & \cdots & 0 & 0 & \cdots & 0 & \cdots & 0 \\ \hline \binom{1}{N-1} & N \cdot M_{AG}^{1,0} & 0 & 0 & 0 & \cdots & 0 & B_{AG}^N & \cdots & 0 & 0 & \cdots & 0 & \cdots & 0 \\ \binom{1}{N-1} & 0 & N \cdot M_{AG}^{N,N+1} & 0 & 0 & \cdots & 0 & 0 & \cdots & B_{AG}^N & 0 & \cdots & 0 & \cdots & 0 \\ \vdots & N \cdot M_{AG}^{1,0} & 0 & 0 & 0 & \cdots & 0 & 0 & \cdots & 0 & B_{AG}^N & \cdots & 0 & \cdots & 0 \\ \hline \binom{1}{N-1} & 0 & 0 & N \cdot M_{AT}^{N,N+1} & 0 & \cdots & 0 & 0 & \cdots & 0 & 0 & \cdots & B_{AT}^N & \cdots & 0 \\ \binom{1}{N-1} & 0 & 0 & 0 & N \cdot M_{AT}^{N,N+1} & \cdots & \vdots & \vdots & \cdots & \vdots & \vdots & \cdots & \vdots & \cdots & \vdots \end{array} \right)
 \end{array}$$

Pros and cons of the Pomo model

1. Pros:

1. bypasses gene tree estimation (easy to combine with DL or DTL models...)
2. does not assume linkage between sites

2. Cons:

1. You don't get gene trees
2. It's not sure what branch lengths mean anymore
3. matrix can become big
4. Only bimorphic sites are considered: 3- or 4-morphic sites need to be handled differently
5. Not sure how it behaves in practice