

```
$ wget http://ssolo.web.elte.hu/lab_slides.pdf
```

```
$ cd ~/workshop_materials
```

```
$ wget http://ssolo.web.elte.hu/lab_data.tgz
```

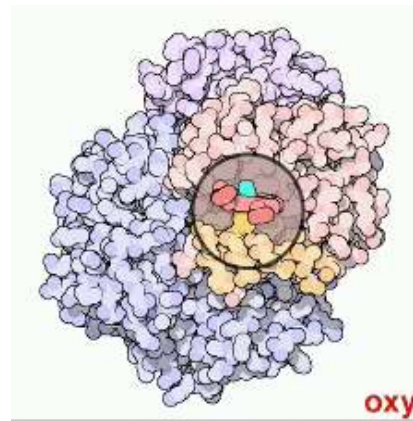
```
$ tar xzf lab_data.tgz
```

```
$ cd lab_data
```

The stories of gene families can be complicated

The story of each gene family consist of a unique series of evolutionary events that often results in a change of copy number and shifts in function.

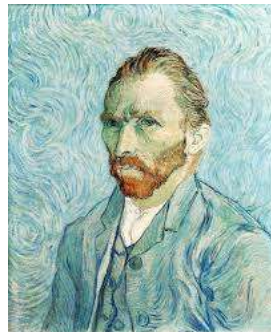
Human hemolglobin is composed of



$2\alpha + 2\beta$ chains.

molecular machine

Human



adult

$+2\beta$ (97%)
 $+2\delta$ (3%)

Cow



adult

$+2\{\beta\delta\}$

Horse



adult and fetus

$+2\{\beta\delta\}$

$2\alpha +$



fetus

$+2\gamma$



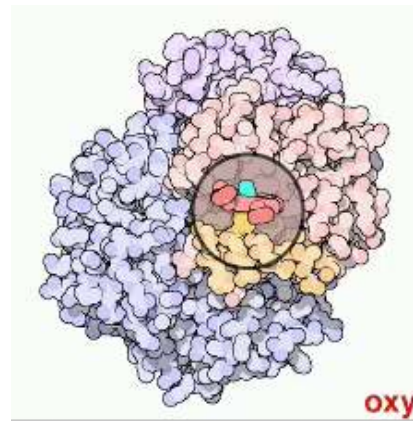
fetus

$+2\gamma$

The stories of gene families can be complicated

The story of each gene family consist of a unique series of evolutionary events that often results in a change of copy number and shifts in function.

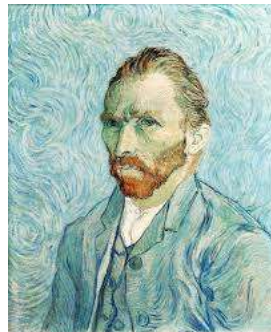
Human hemolglobin is composed of



$2\alpha + 2\beta$ chains.

molecular machine

Human



adult

$+2\beta$ (97%)
 $+2\delta$ (3%)

Cow



adult

$+2\{\beta\delta\}$

Horse



adult and fetus

$+2\{\beta\delta\}$

$2\alpha +$



fetus

$+2\gamma$

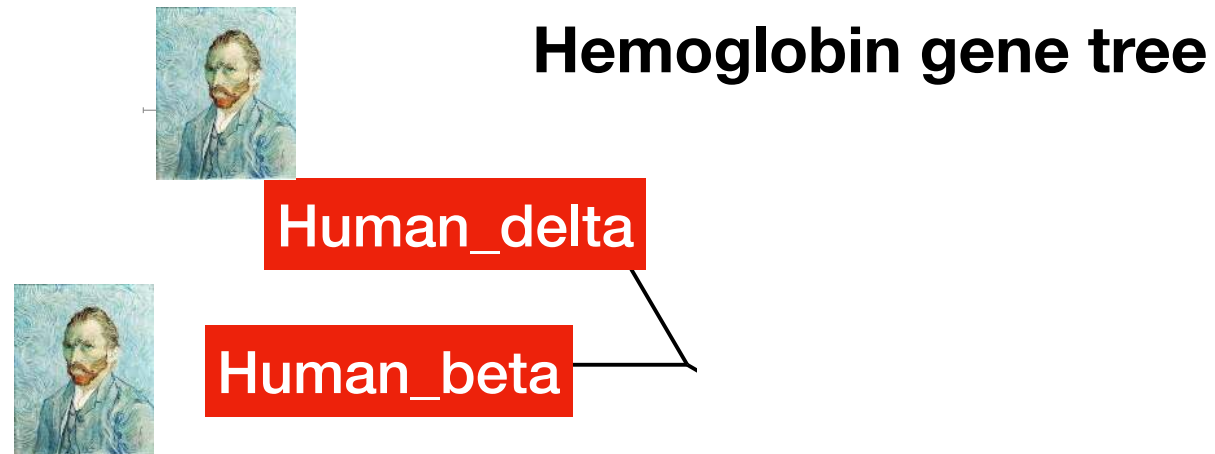


fetus

$+2\gamma$

Hemoglobin/

```
$ cd ~/workshop_materials/lab_data/Hemoglobin
```



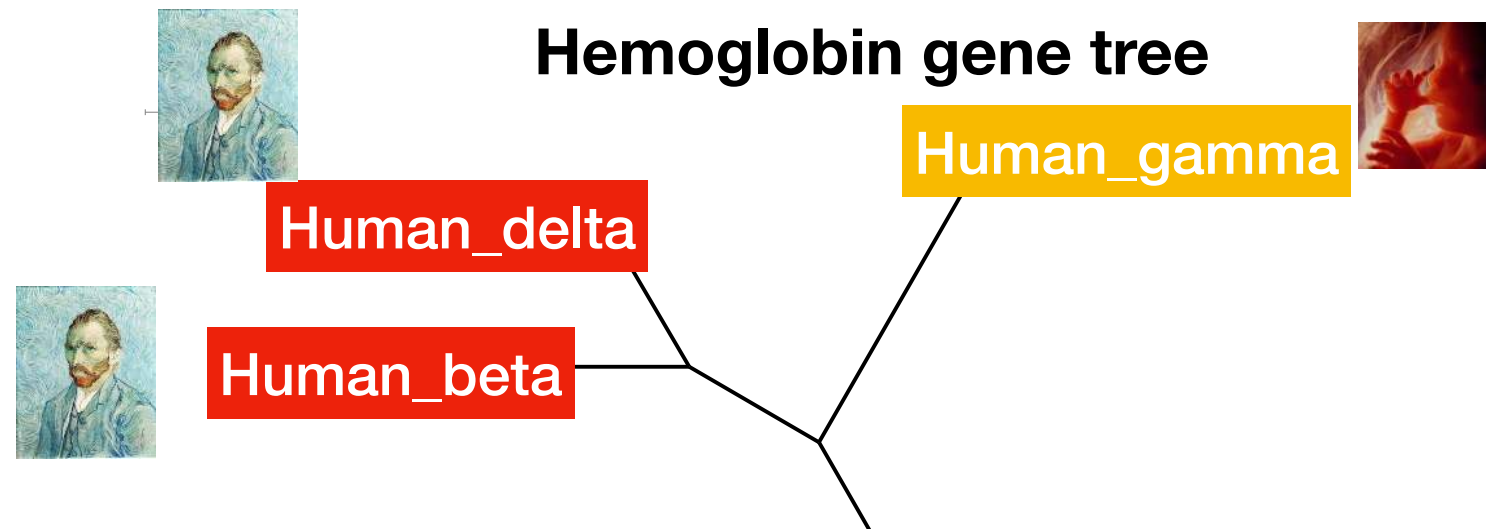
Hemoglobin.tree

(Human_beta:1,Human_delta:1):1,



Hemoglobin/

```
$ cd ~/workshop_materials/lab_data/Hemoglobin
```



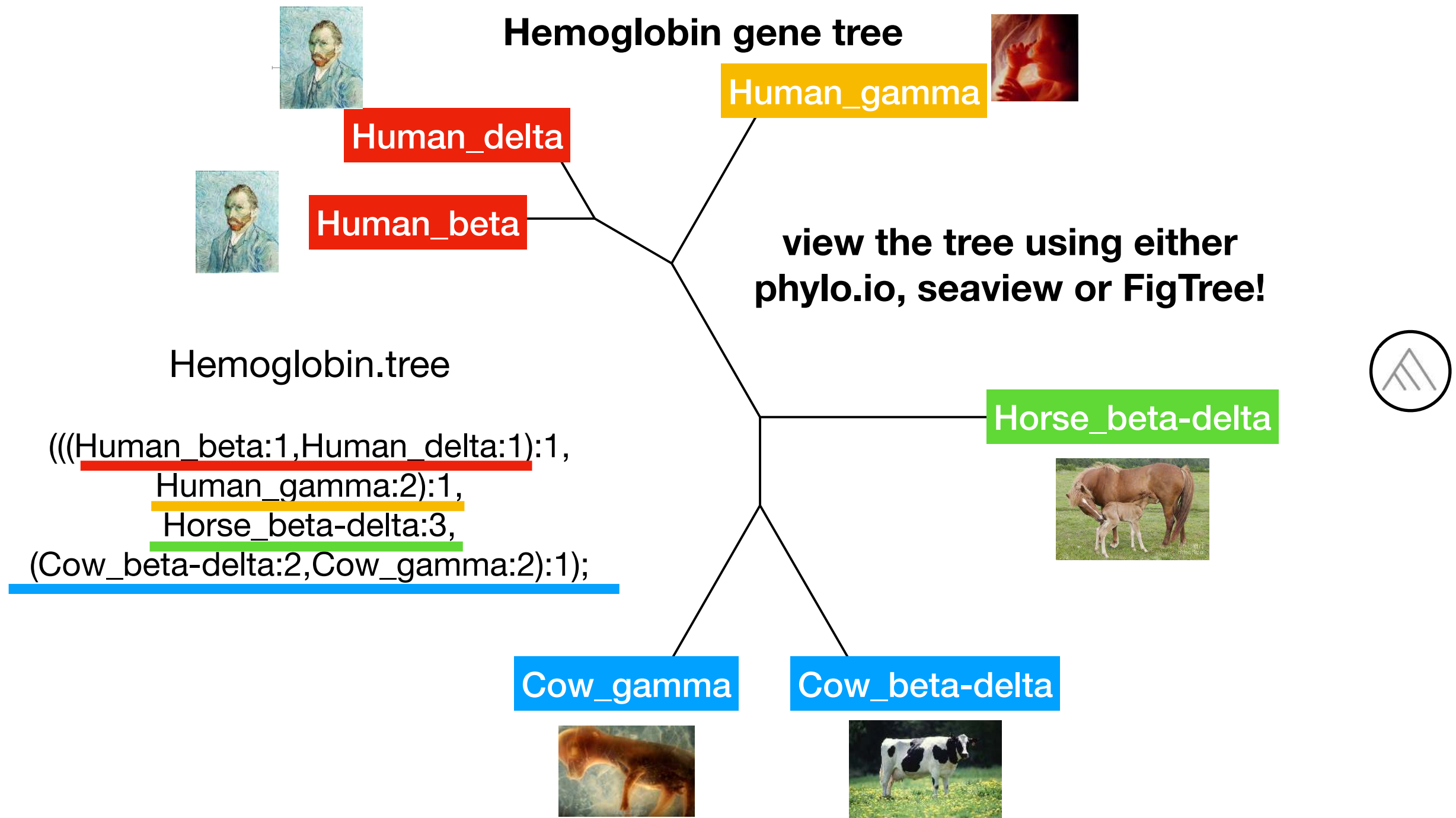
Hemoglobin.tree

```
((Human_beta:1,Human_delta:1):1,  
Human_gamma:2):1,
```



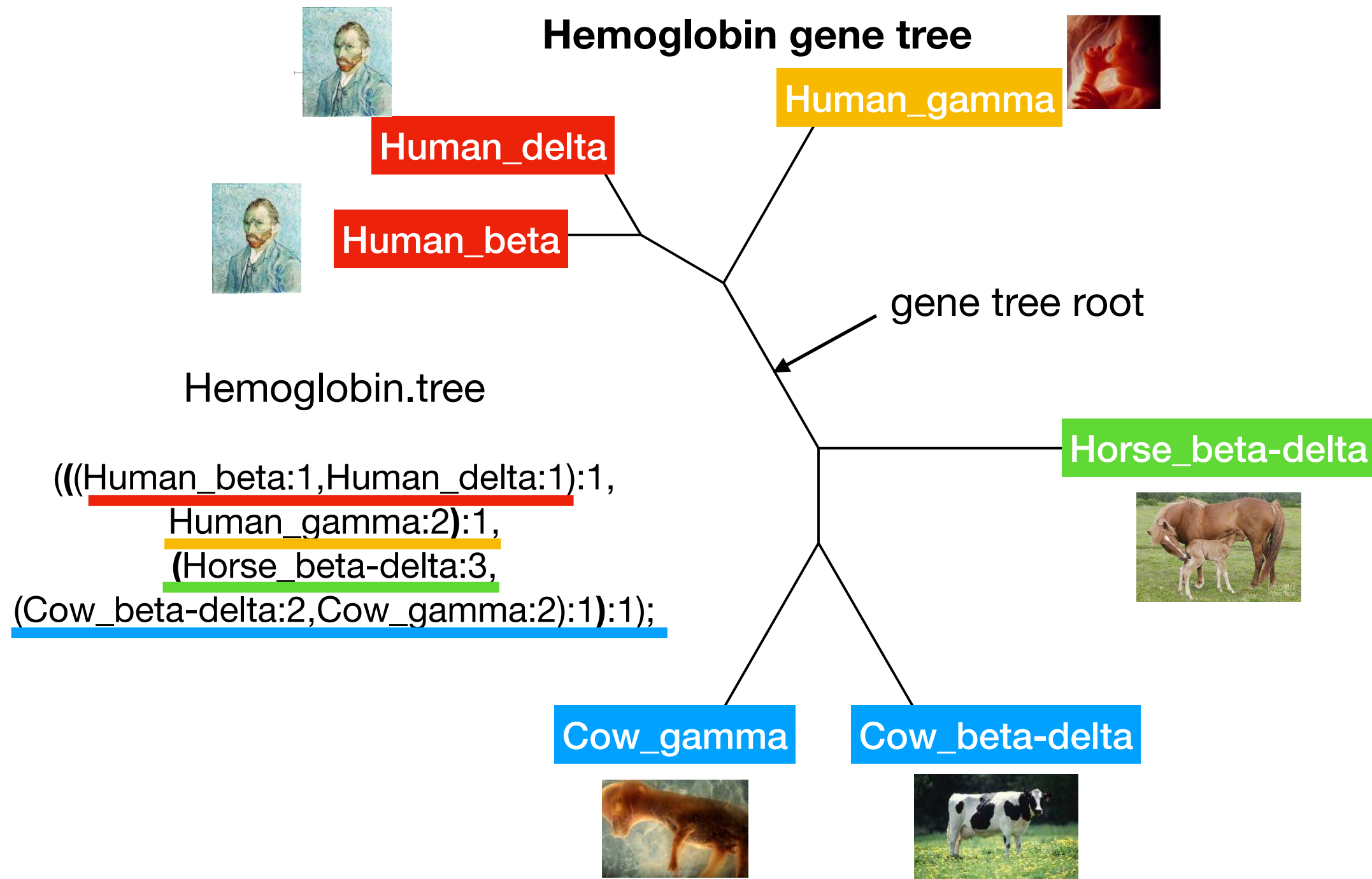
Hemoglobin/

```
$ cd ~/workshop_materials/lab_data/Hemoglobin
```



Hemoglobin/

```
$ cd ~/workshop_materials/lab_data/Hemoglobin
```

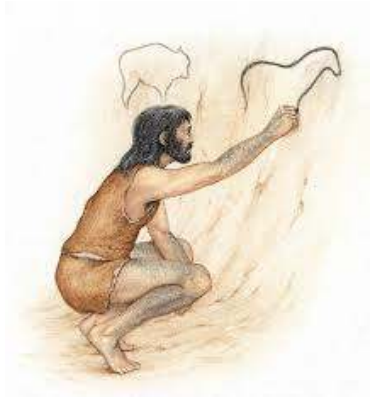
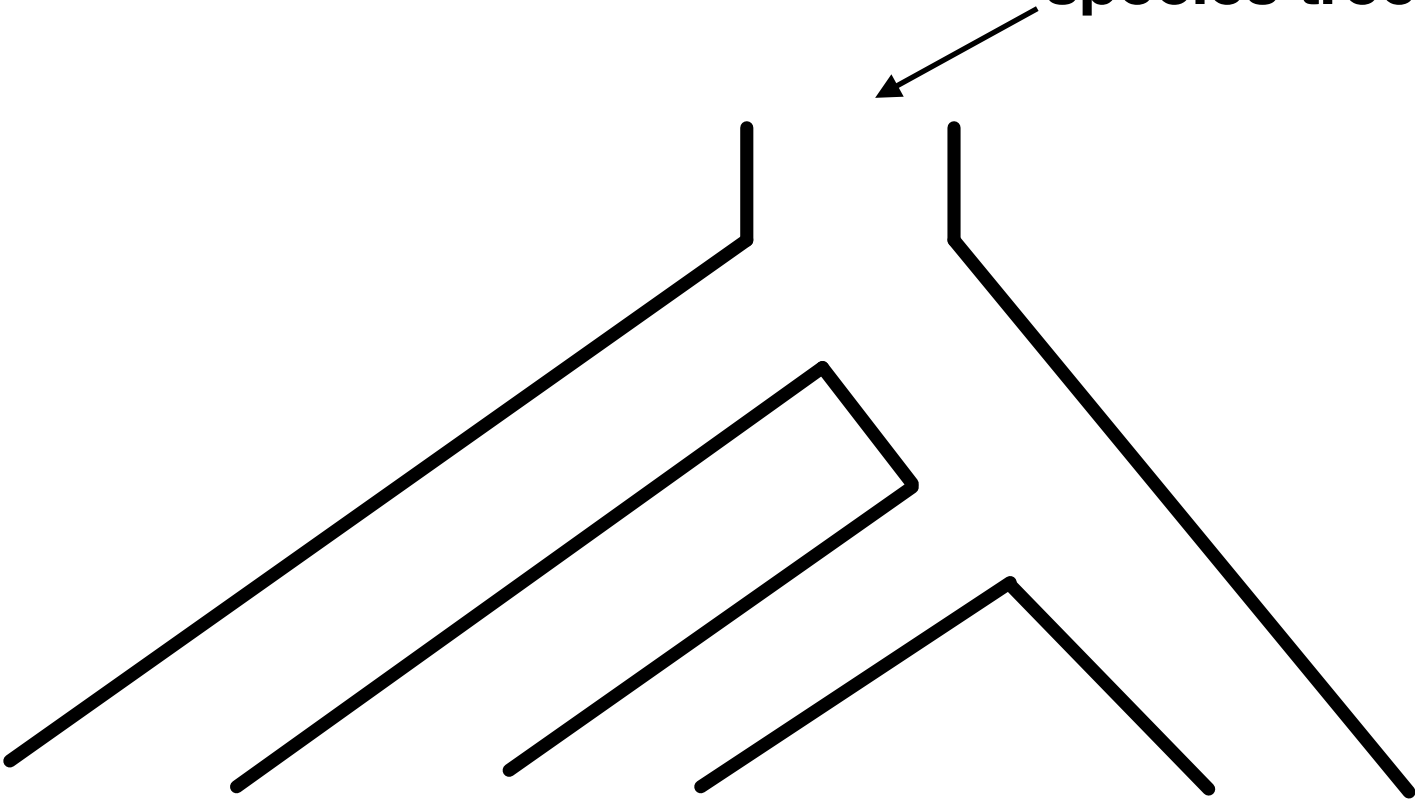


Species tree

HuHoCo.tree

(Human:2,(Horse:1,Cow:1):1);

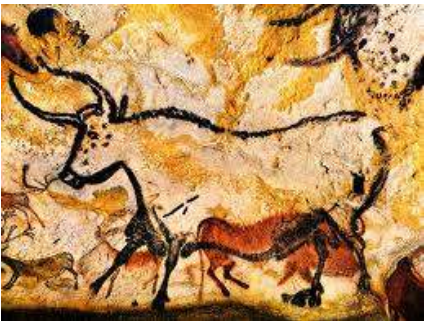
species tree root



Human



Horse

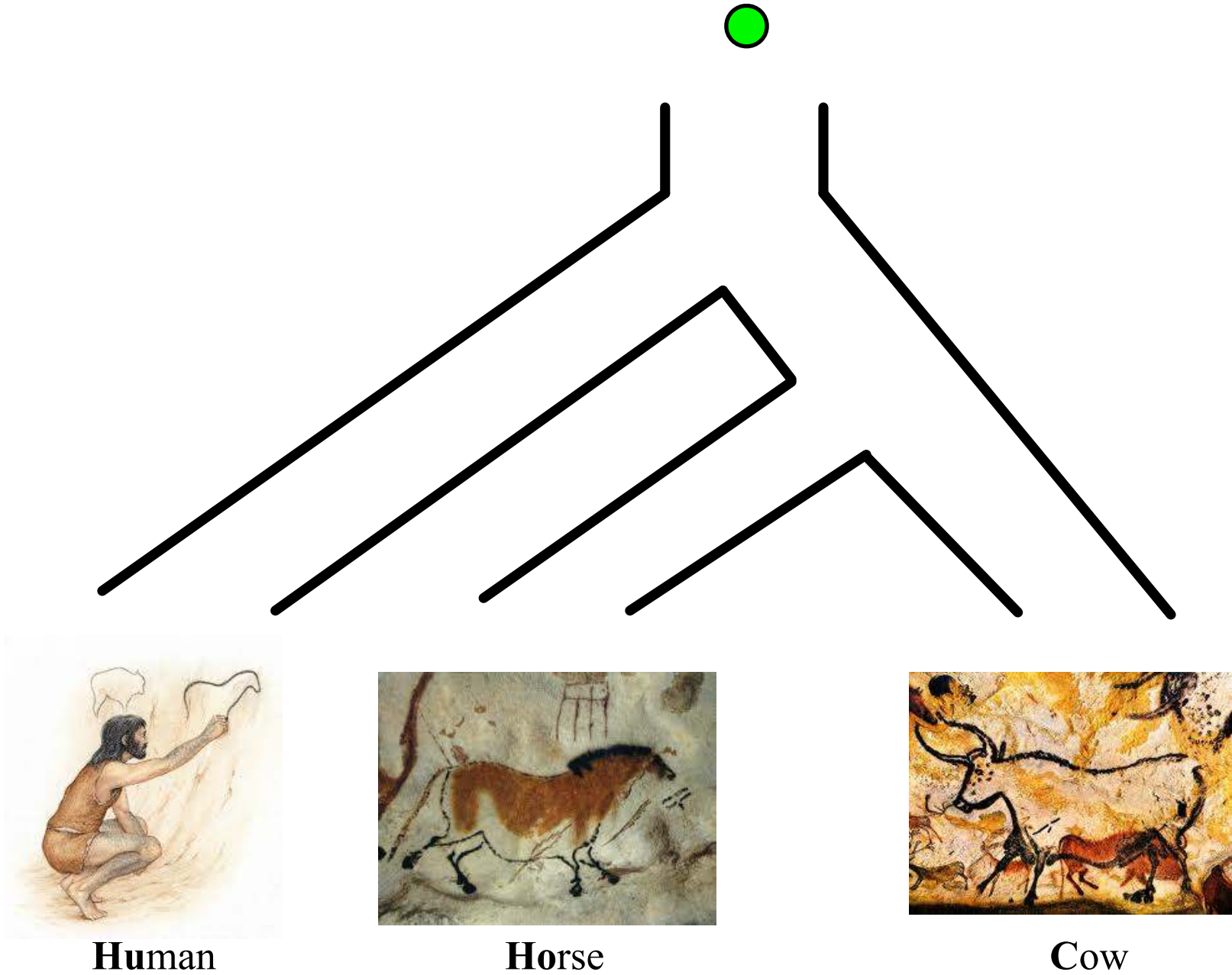


Cow

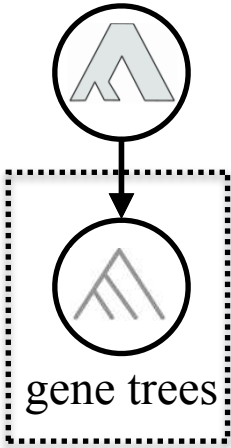
species

Hemoglobin/

DL along S



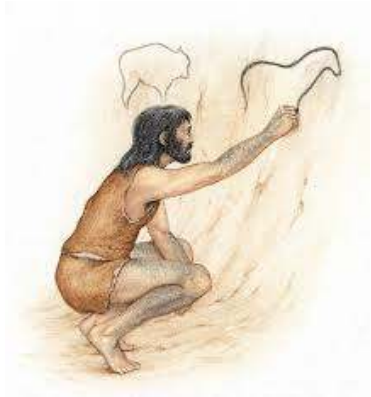
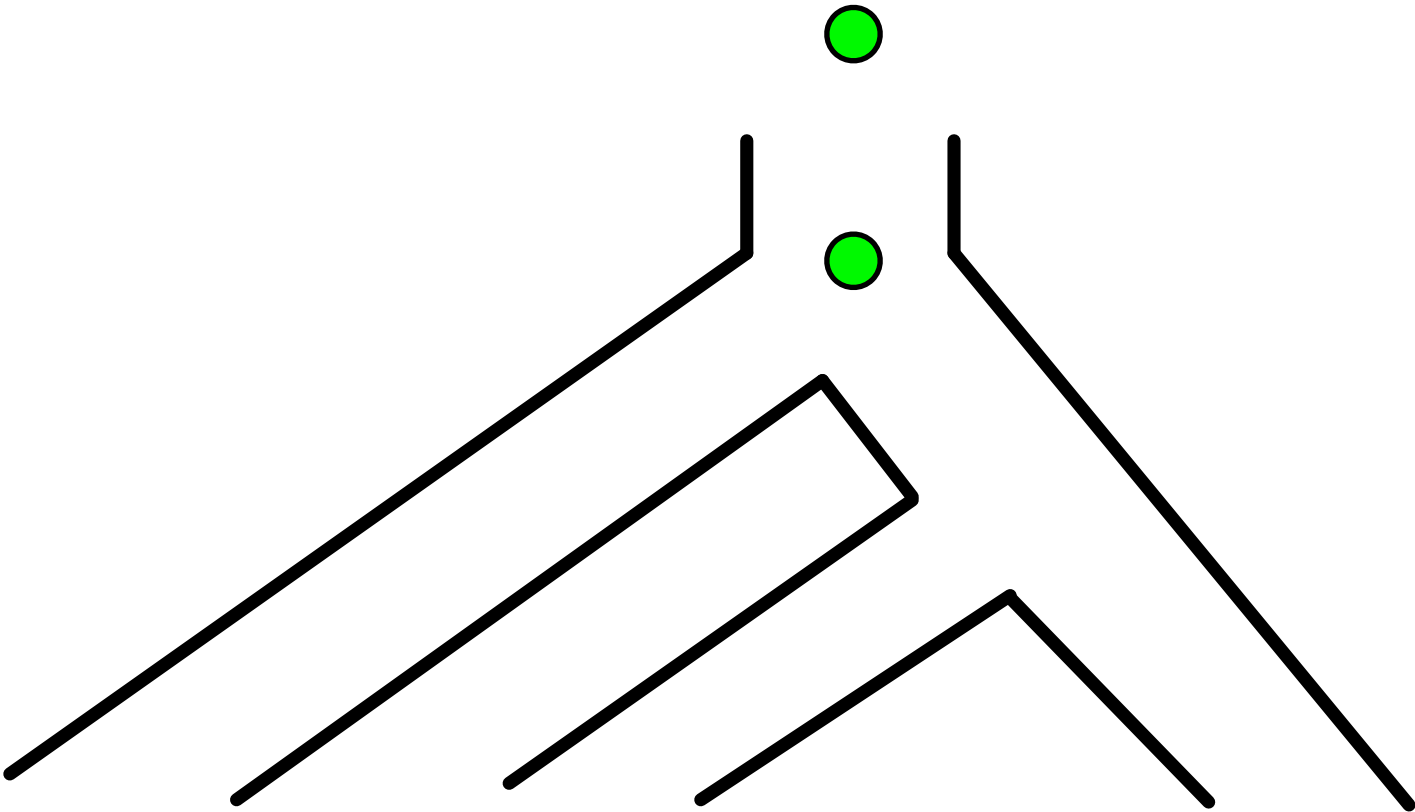
DL
species tree



species

Hemoglobin/

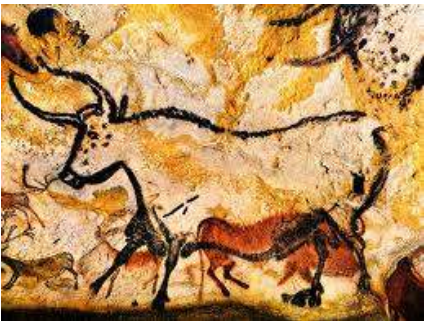
DL along S



Human

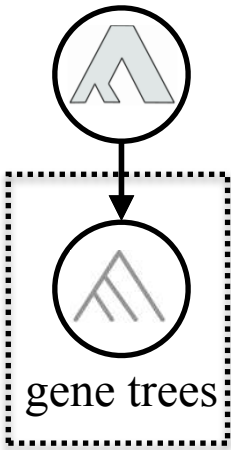


Horse



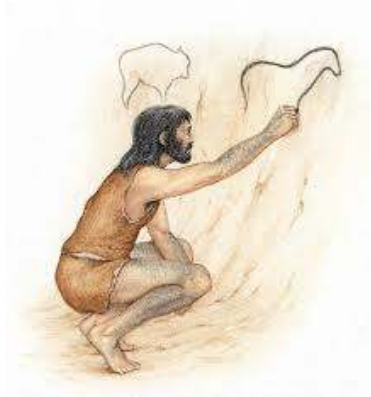
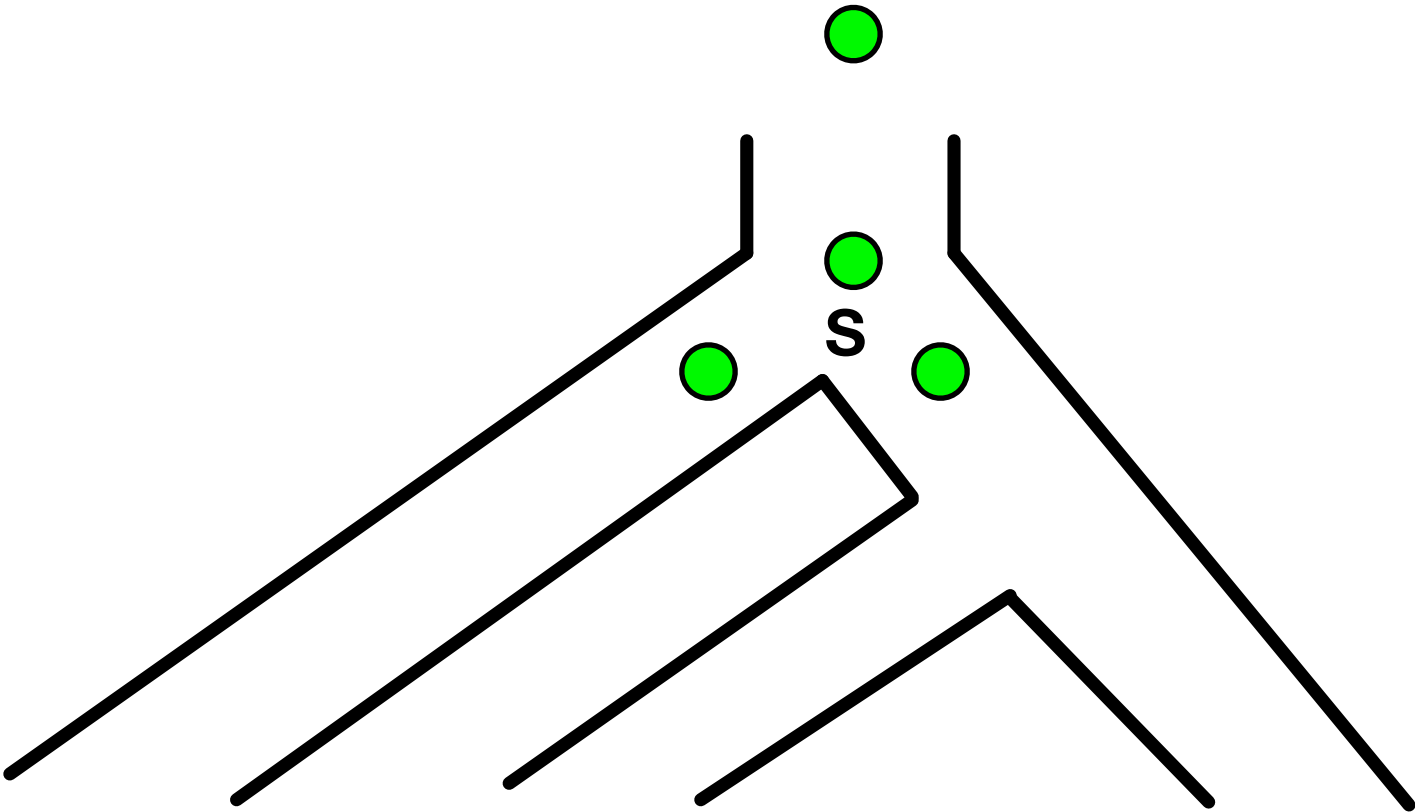
Cow

DL
species tree



species

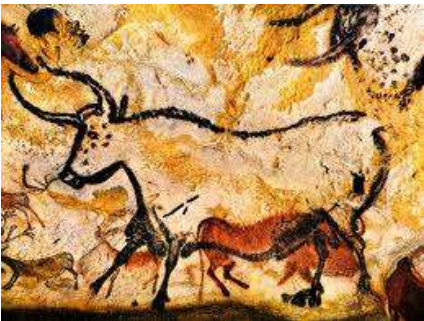
DL along S



Human

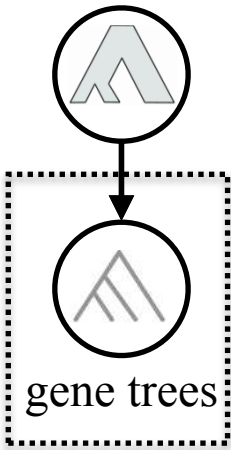


Horse



Cow

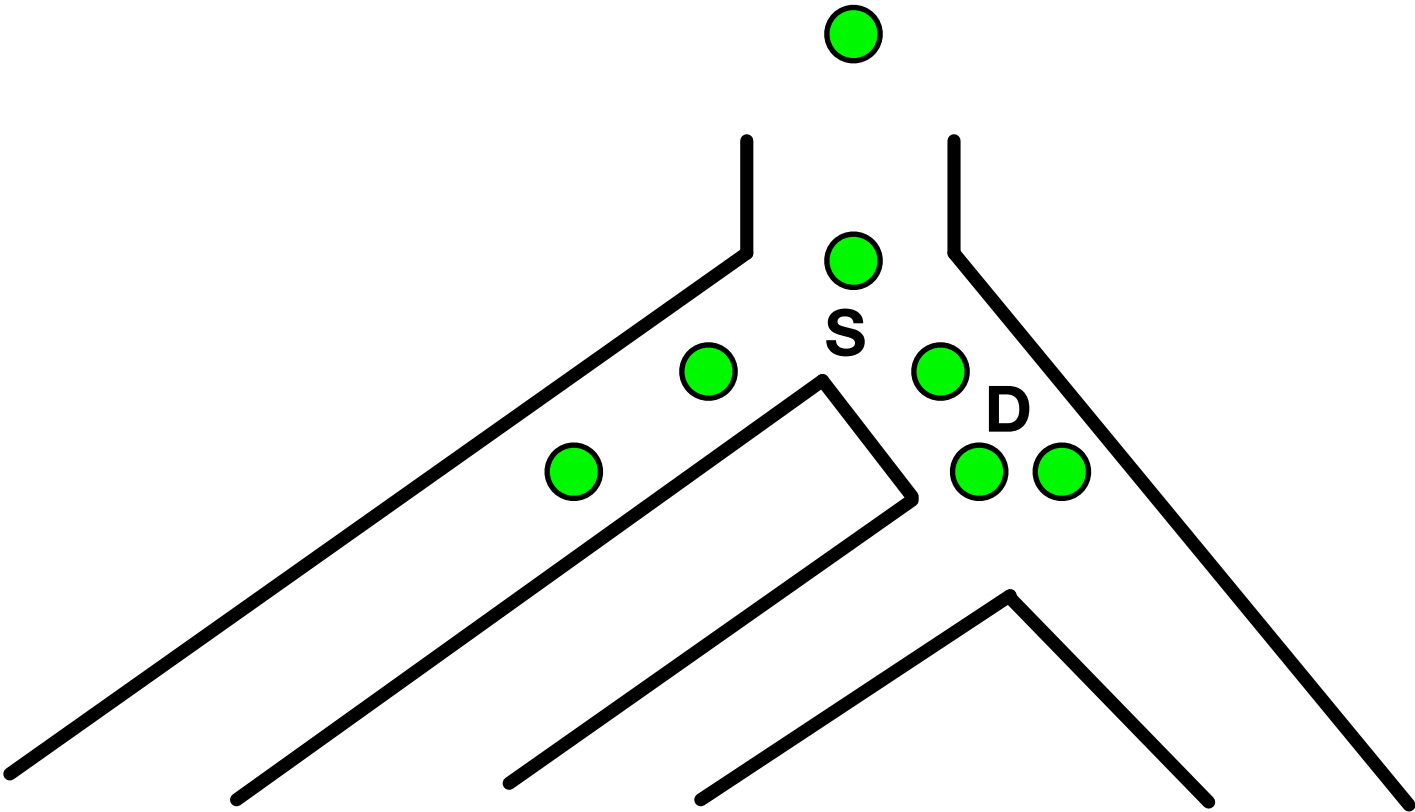
DL
species tree



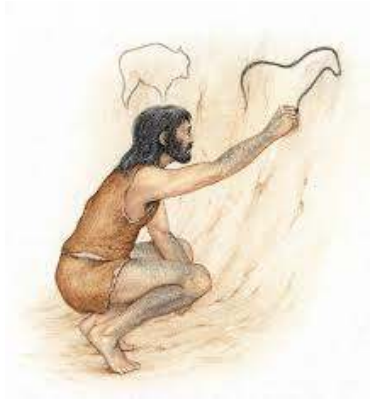
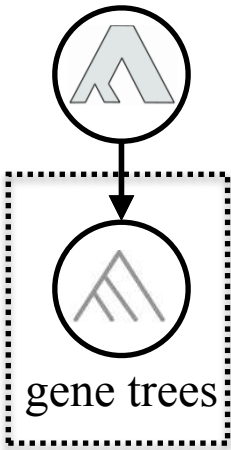
species

Hemoglobin/

DL along S



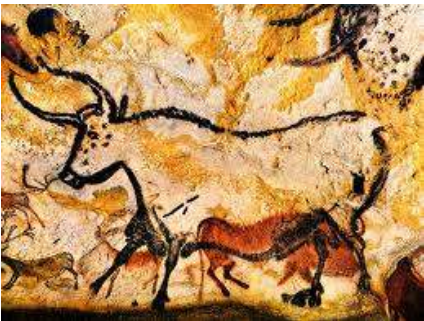
DL
species tree



Human



Horse

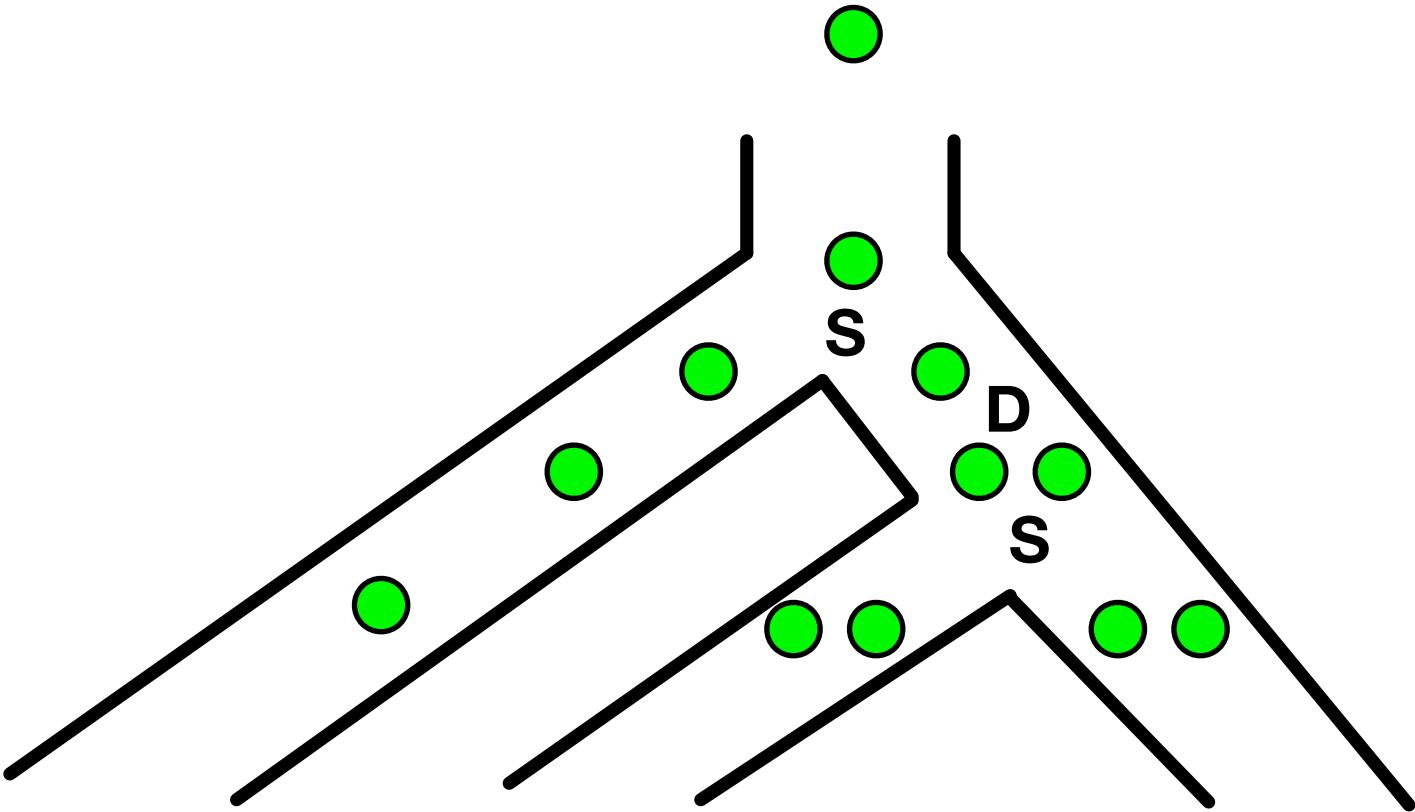


Cow

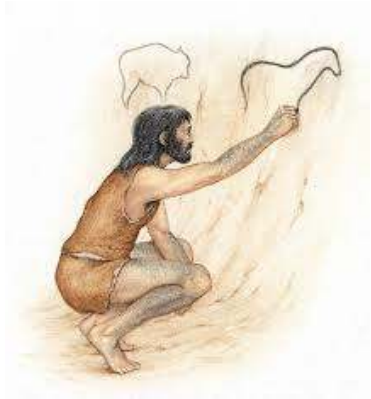
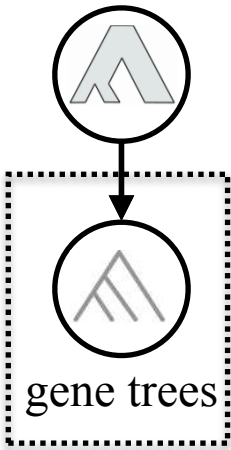
species

Hemoglobin/

DL along S



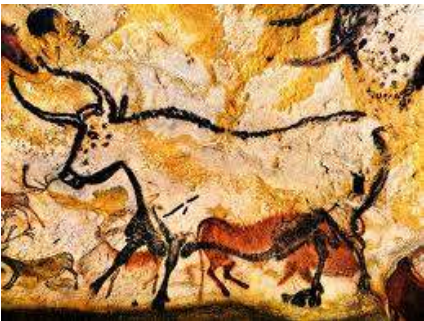
DL
species tree



Human



Horse

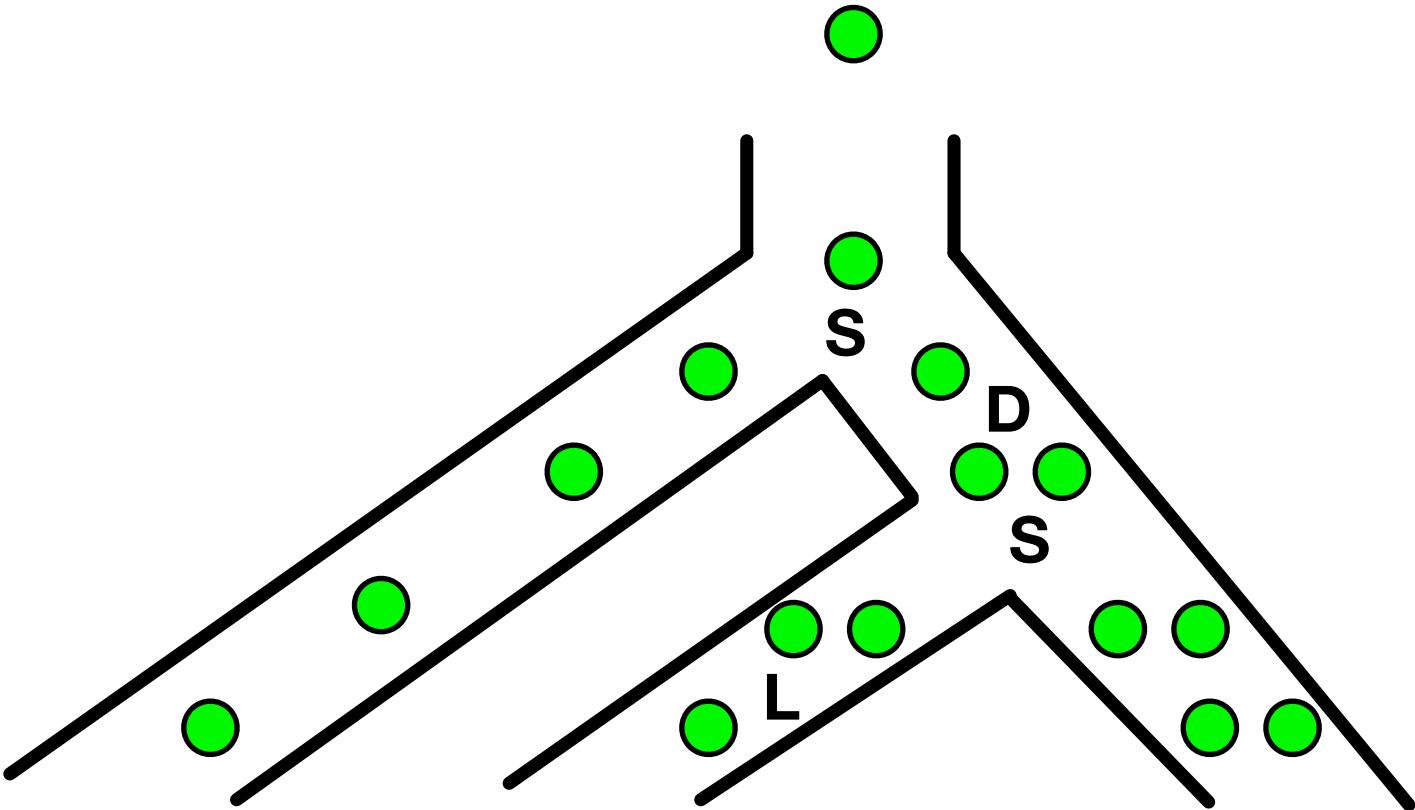


Cow

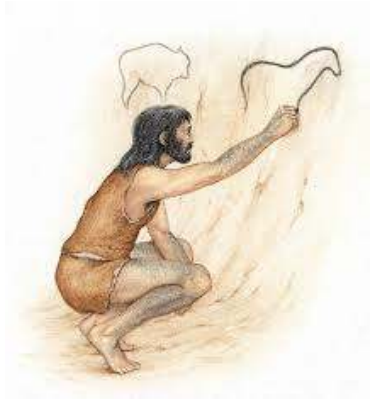
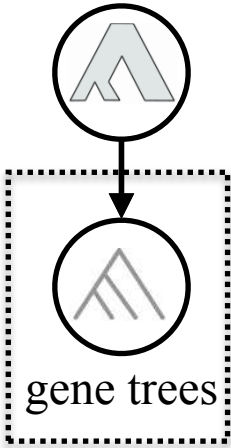
species

Hemoglobin/

DL along S



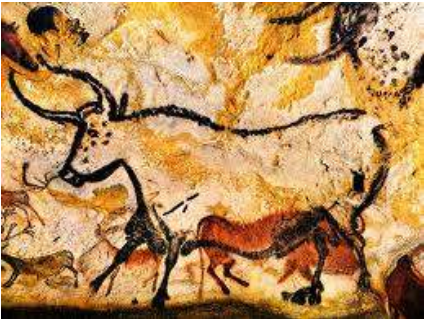
DL
species tree



Human



Horse

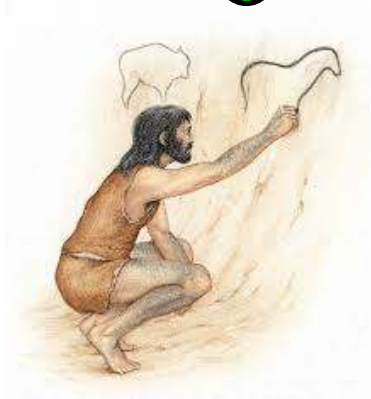
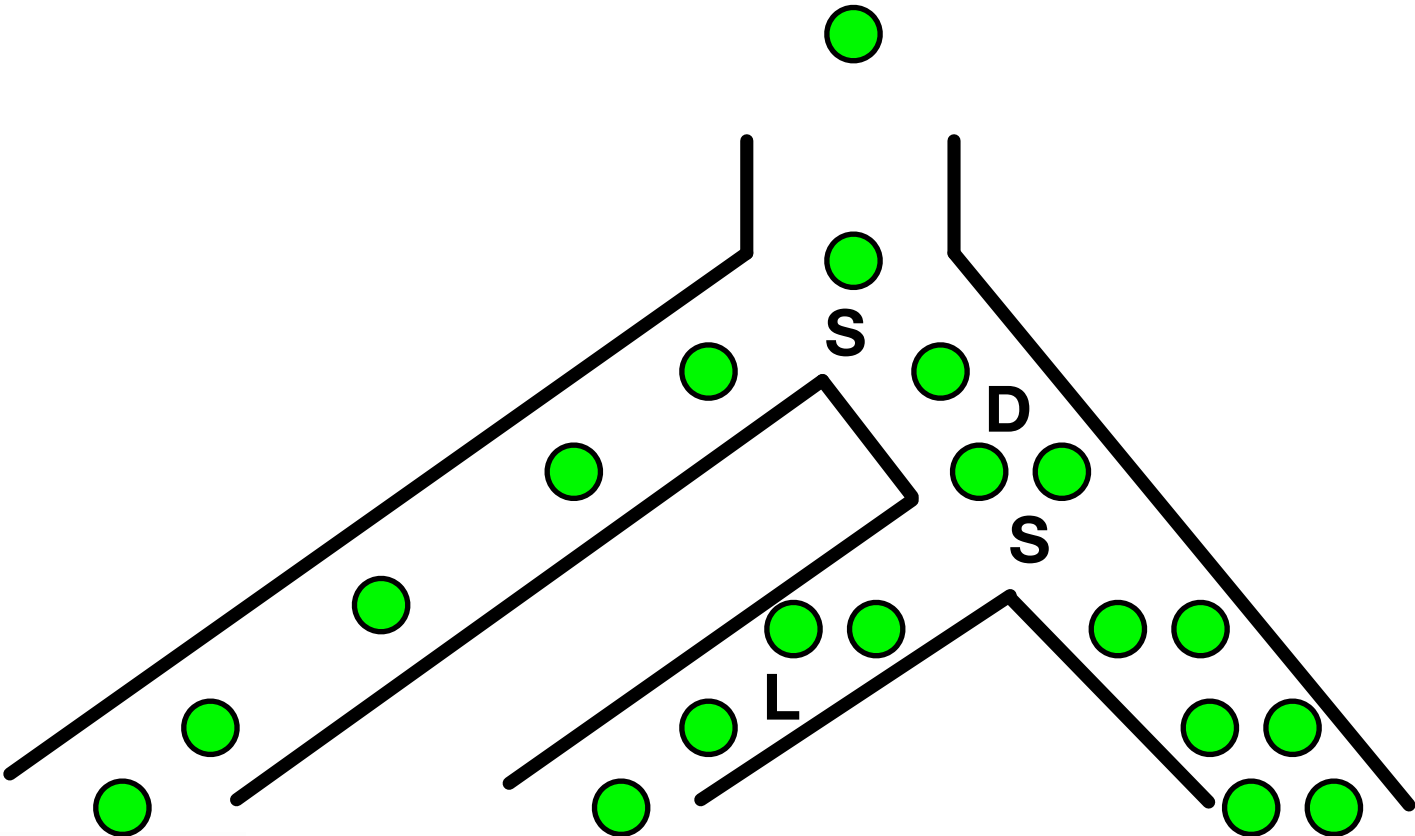


Cow

species

Hemoglobin/

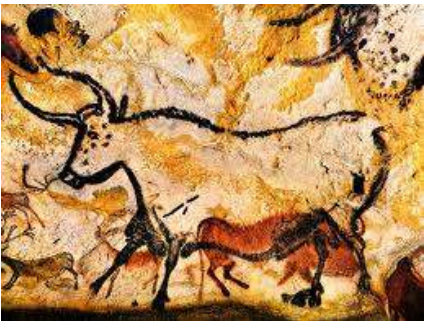
DL along S



Human

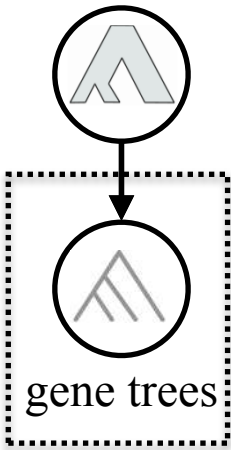


Horse



Cow

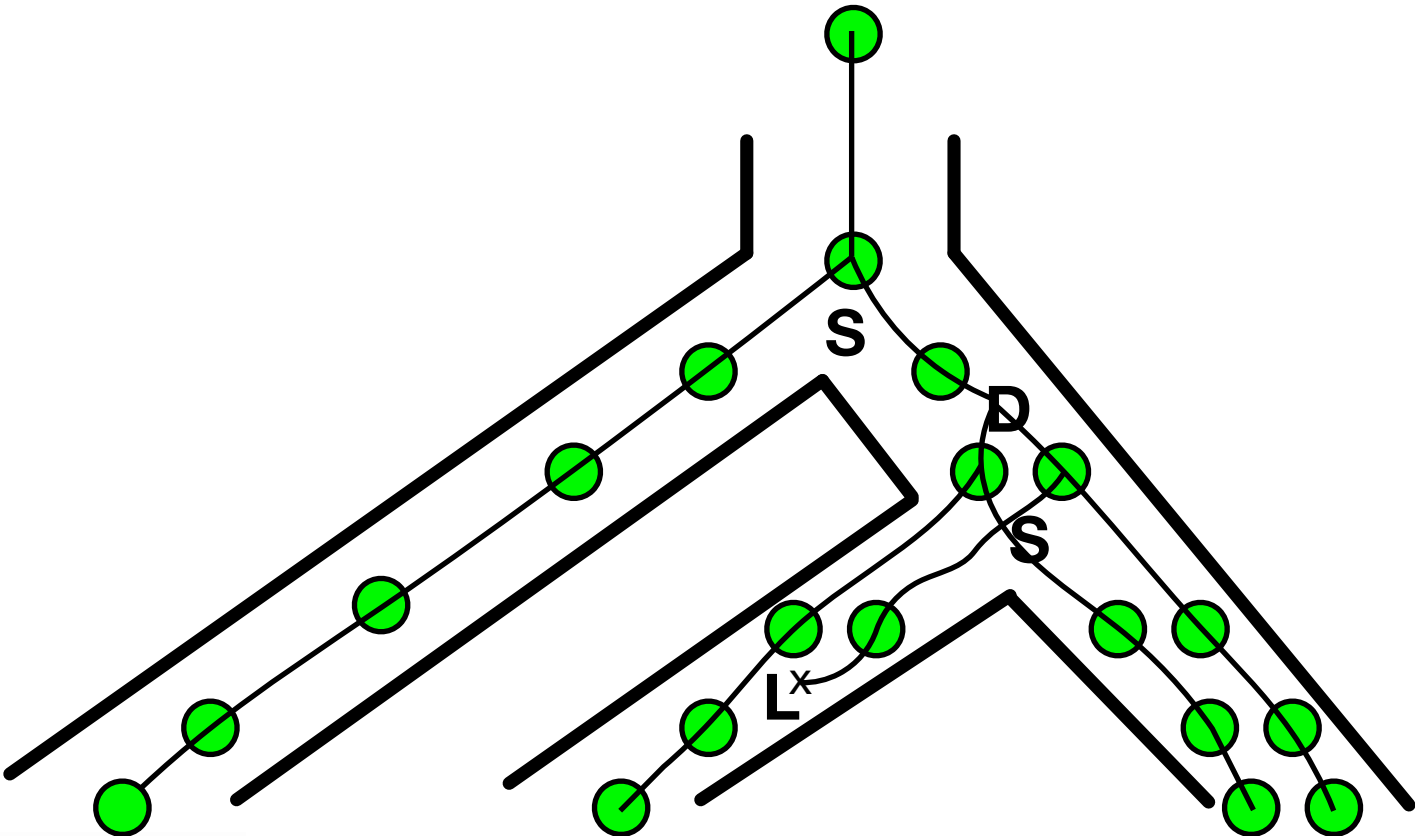
DL
species tree



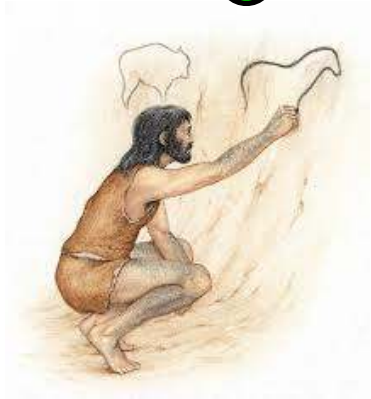
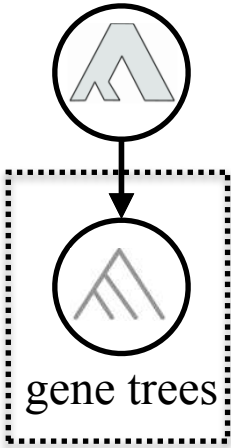
species

Hemoglobin/

DL along S



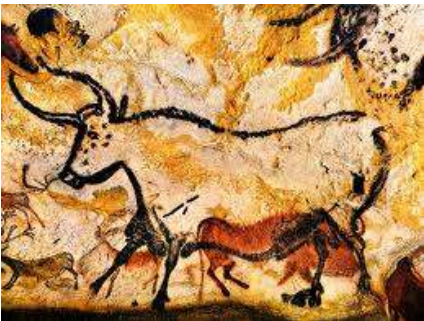
DL
species tree



Human



Horse



Cow

species

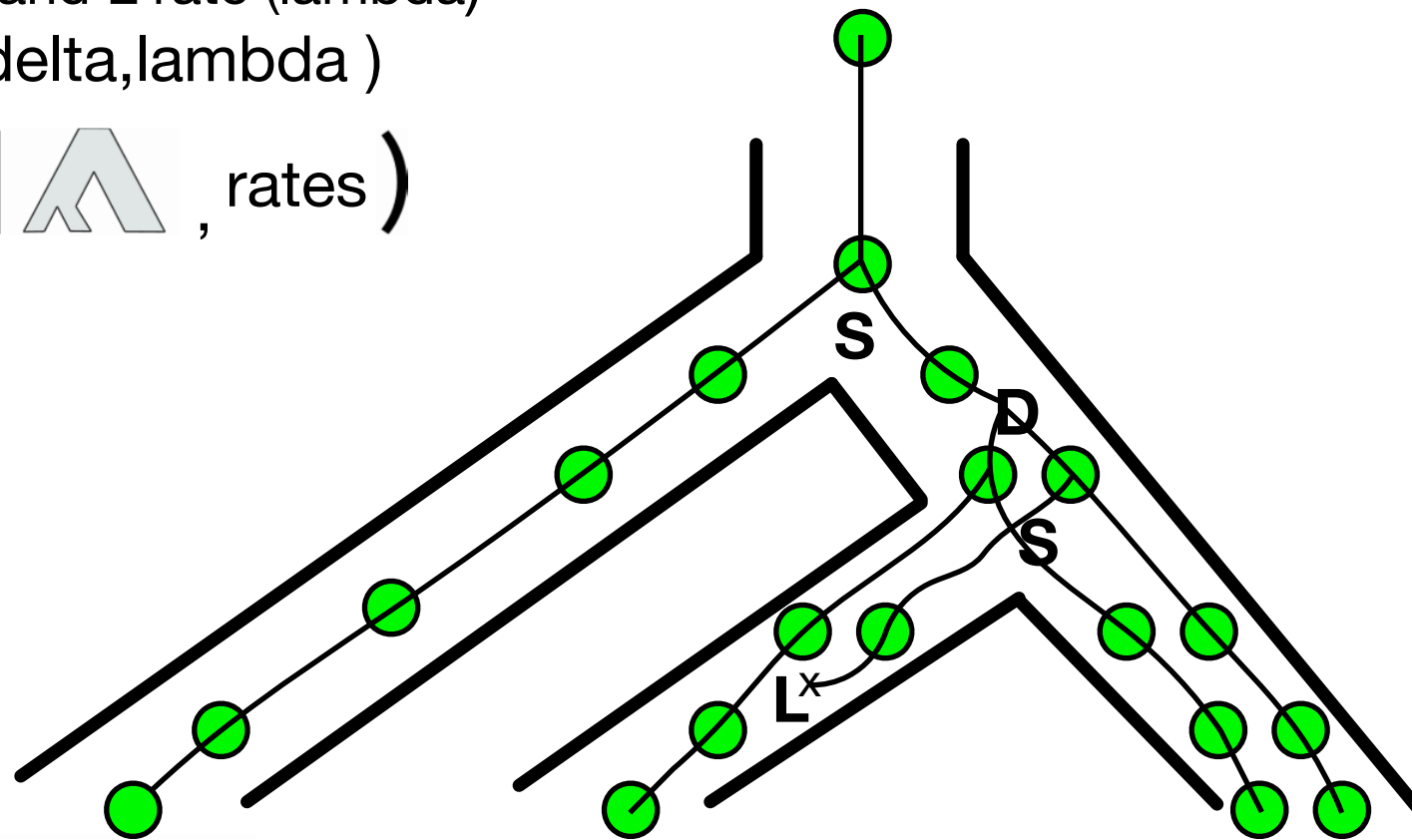
Hemoglobin/

DL along S

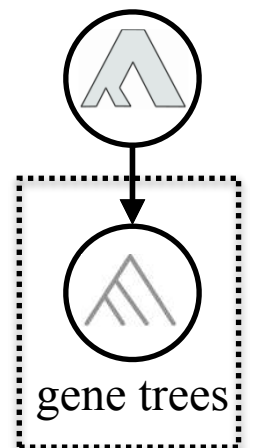
Given G and S we want to calculate:

1. D rate (delta) and L rate (lambda)
2. $P(G \mid S, \text{delta}, \text{lambda})$

$$P(\text{gene tree} \mid \text{species tree}, \text{rates})$$



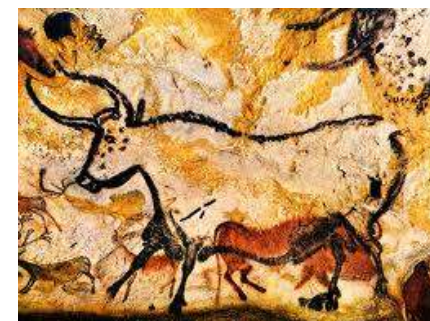
DL
species tree



Human



Horse



Cow

species

Hemoglobin/

ALEobserve Hemoglobin.tree

ALEml HuHoCo.tree Hemoglobin.tree.ale tau=0 sample=10

less HuHoCo.tree_Hemoglobin.tree.ale.ml_rec

#ALEml using ALE v0.5 by Szollosi GJ et al.; ssolo@elte.hu; CC BY-SA 3.0;

S: (Human:1,(Horse:0.5,Cow:0.5)1:0.5)2; ← species tree

Input ale from: Hemoglobin.tree.ale
>logl: -4.45543 ← ML log-likelihood (tau=0)

	rate of Duplications	Transfers	Losses
ML	0.76128 0	1e-10	

← ML rates (tau=0)

10 reconciled G-s:

((Human_gamma:2,(Human_beta:1,Human_delta:1)D@0|0|Human:1)D@0|0.05|Human:2,(Horse_beta-delta:3,(Cow_gamma:2,Cow_beta-delta:2)D@0|0.3|Cow:1).1:2).2:0;
((Human_gamma:2,(Human_beta:1,Human_delta:1)D@1|0.5|Human:1)D@1|0.55|Human:2,(Horse_beta-delta:3,(Cow_gamma:2,Cow_beta-delta:2)D@0|0.2|Cow:1).1:2).2:0;
((Human_gamma:2,(Human_beta:1,Human_delta:1)D@0|0.1|Human:1)D@0|0.4|Human:2,(Horse_beta-delta:3,(Cow_gamma:2,Cow_beta-delta:2)D@0|0.4|Cow:1).1:2).2:0;
((Human_gamma:2,(Human_beta:1,Human_delta:1)D@0|0.15|Human:1)D@0|0.25|Human:2,(Horse_beta-delta:3,(Cow_gamma:2,Cow_beta-delta:2)D@0|0.2|Cow:1).1:2).2:0;
((Human_gamma:2,(Human_beta:1,Human_delta:1)D@0|0.25|Human:1)D@1|0.75|Human:2,(Horse_beta-delta:3,(Cow_gamma:2,Cow_beta-delta:2)D@0|0.4|Cow:1).1:2).2:0;
((Human_gamma:2,(Human_beta:1,Human_delta:1)D@0|0.1|Human:1)D@0|0.25|Human:2,(Horse_beta-delta:3,(Cow_gamma:2,Cow_beta-delta:2)D@0|0.2|Cow:1).1:2).2:0;
((Human_gamma:2,(Human_beta:1,Human_delta:1)D@0|0.05|Human:1)D@0|0.3|Human:2,(Horse_beta-delta:3,(Cow_gamma:2,Cow_beta-delta:2)D@0|0.05|Cow:1).1:2).2:0;
((Human_gamma:2,(Human_beta:1,Human_delta:1)D@0|0.2|Human:1)D@1|0.55|Human:2,(Horse_beta-delta:3,(Cow_gamma:2,Cow_beta-delta:2)D@0|0.2|Cow:1).1:2).2:0;
((Human_gamma:2,(Human_beta:1,Human_delta:1)D@0|0.35|Human:1)D@1|0.5|Human:2,(Horse_beta-delta:3,(Cow_gamma:2,Cow_beta-delta:2)D@0|0.05|Cow:1).1:2).2:0;
((Human_gamma:2,(Human_beta:1,Human_delta:1)D@0|0.15|Human:1)D@0|0.3|Human:2,(Horse_beta-delta:3,(Cow_gamma:2,Cow_beta-delta:2)D@0|0.15|Cow:1).1:2).2:0;

# of	Duplications	Transfers	Losses	Speciations
Total	3	0	0	2

# of	Duplications	Transfers	Losses	copies	
S_terminal_branch	Cow	1	0	0	2
S_terminal_branch	Horse	0	0	0	1
S_terminal_branch	Human	2	0	0	3
S_internal_branch	1	0	0	0	1
S_internal_branch	2	0	0	0	1

HuHoCo.tree_Hemoglobin.tree.ale.ml_rec (END)

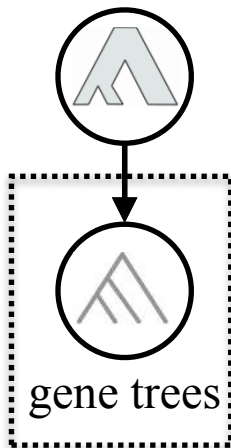
mean number of events
per branch in sampled
reconciliations

10 random reconciled gene trees
sampled according to joint likelihood

$$\sum_{\text{gene trees}} P(\text{events} | \text{gene tree}) P(\text{gene tree} | \text{species tree})$$

sum over reconciled
gene trees

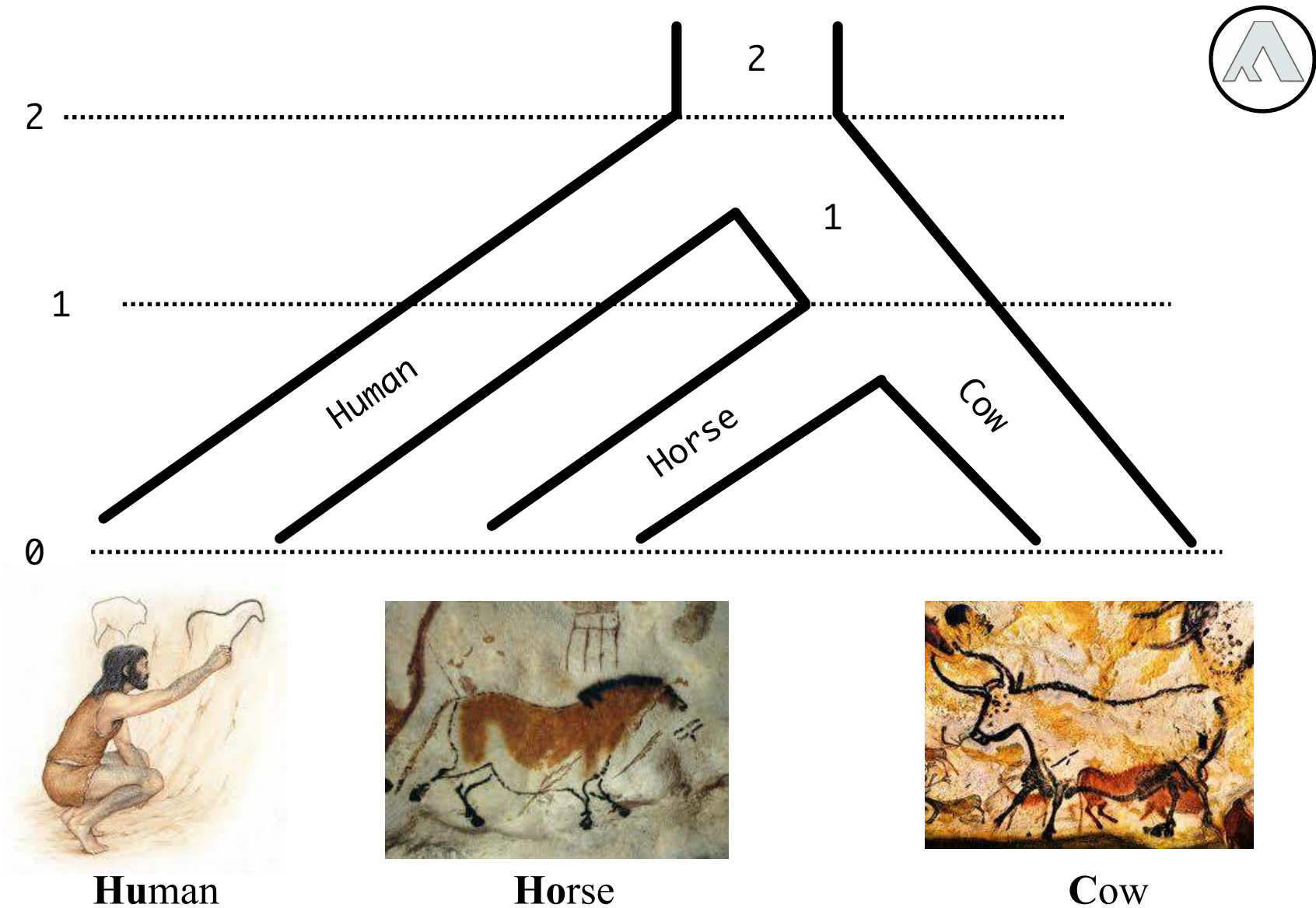
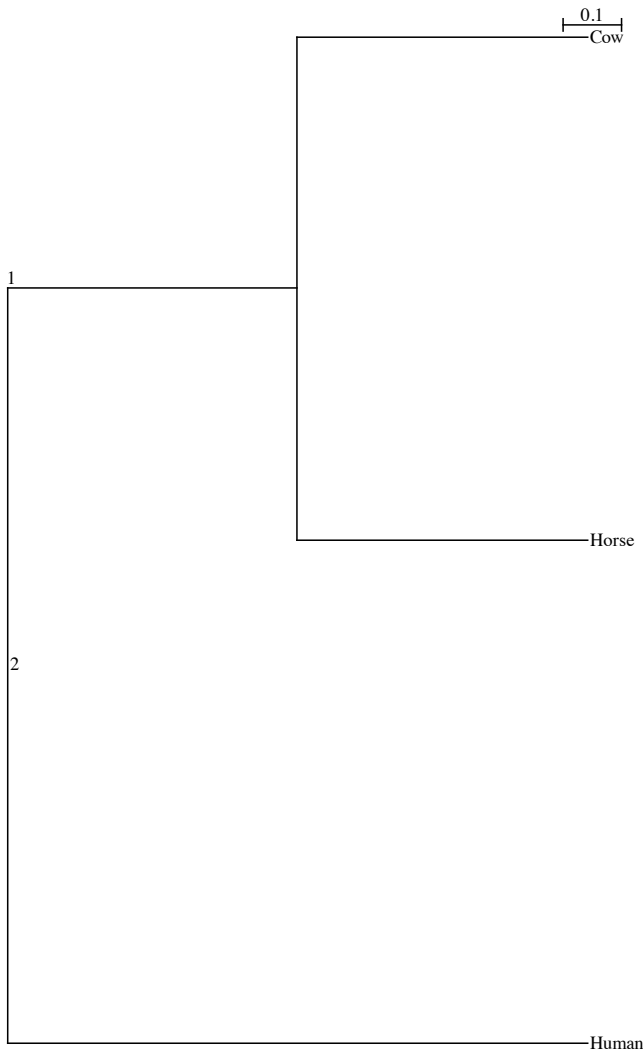
DL
species tree



Hemoglobin/

ALEml HuHoCo.tree Hemoglobin.tree.ale tau=0 sample=10

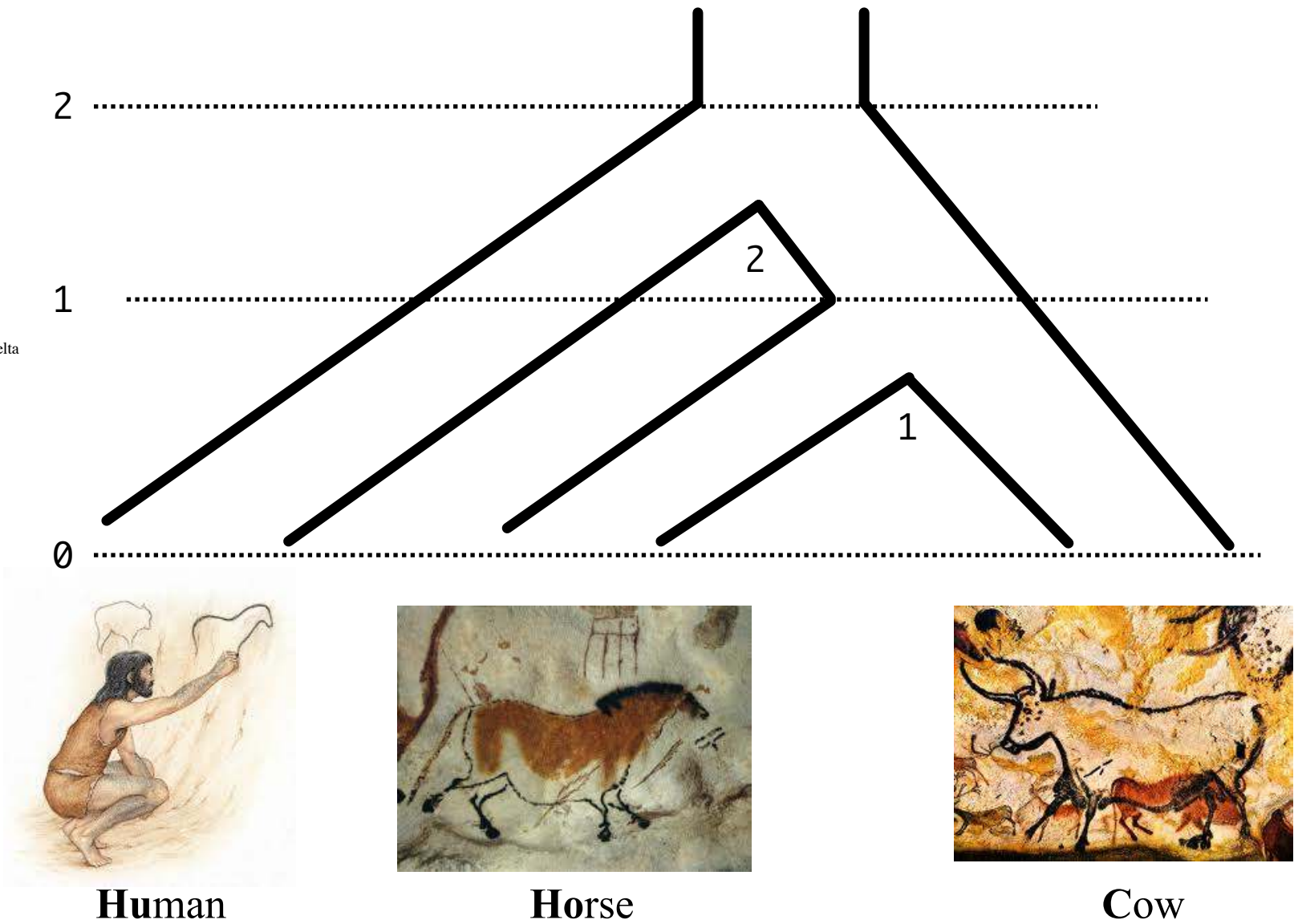
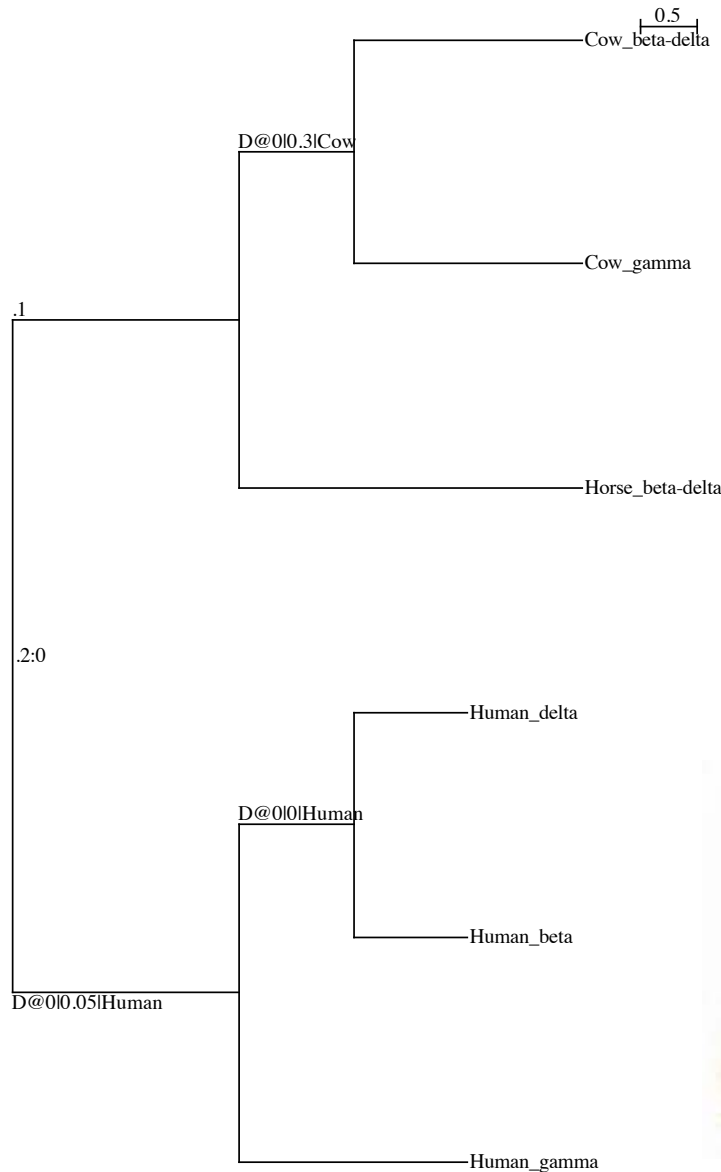
S: (Human:1,(Horse:0.5,Cow:0.5)1:0.5)2;



Hemoglobin/

```
((Human_gamma:2,(Human_beta:1,Human_delta:1)D@0|0.01|Human:1)D@0|0.05|Human:2,(Horse_beta-delta:3,(Cow_gamma:2,Cow_beta-delta:2)D@0|0.3|Cow:1).1:2).2:0;
```

ALEml HuHoCo.tree Hemoglobin.tree.ale tau=0 sample=10

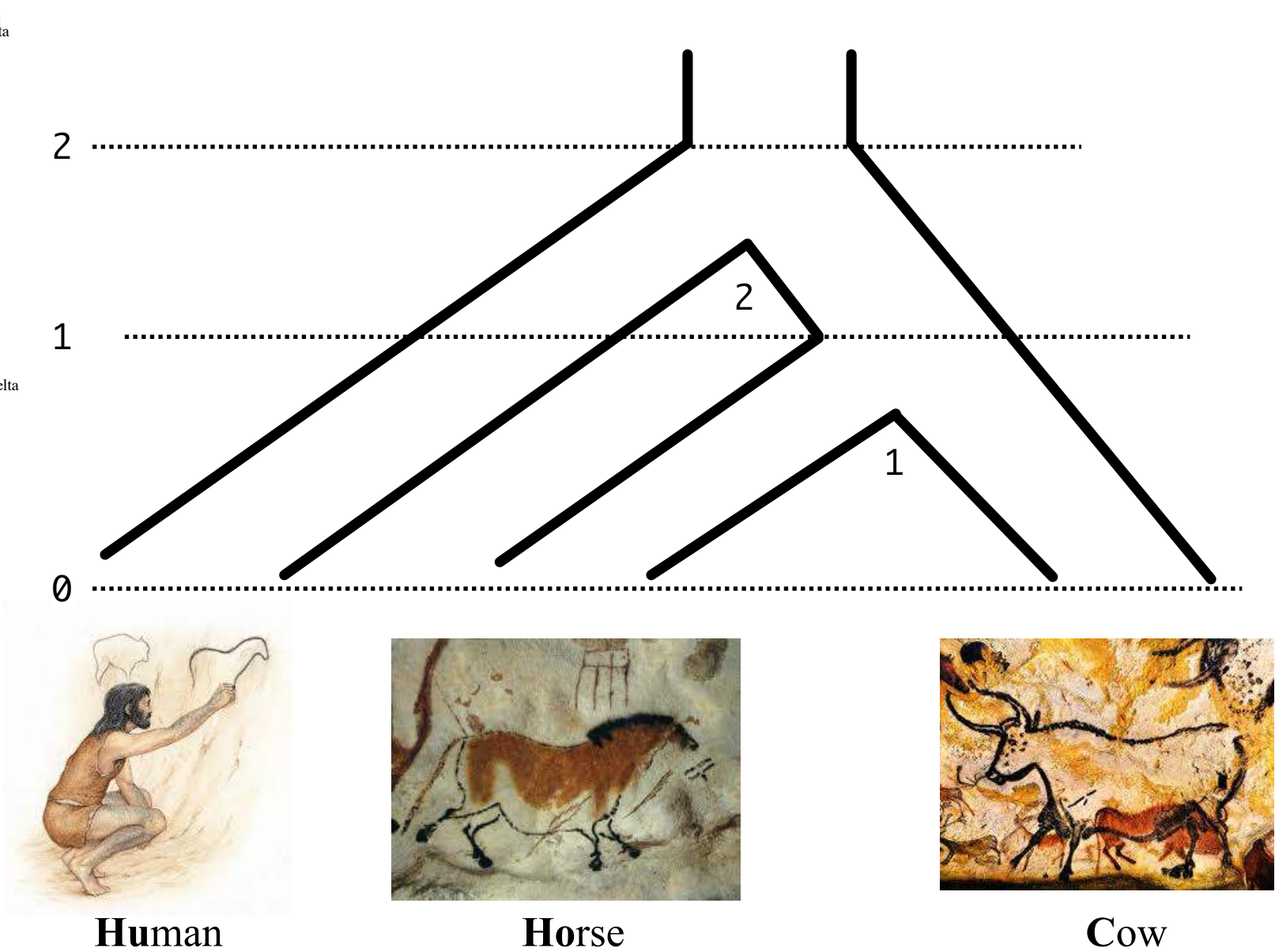
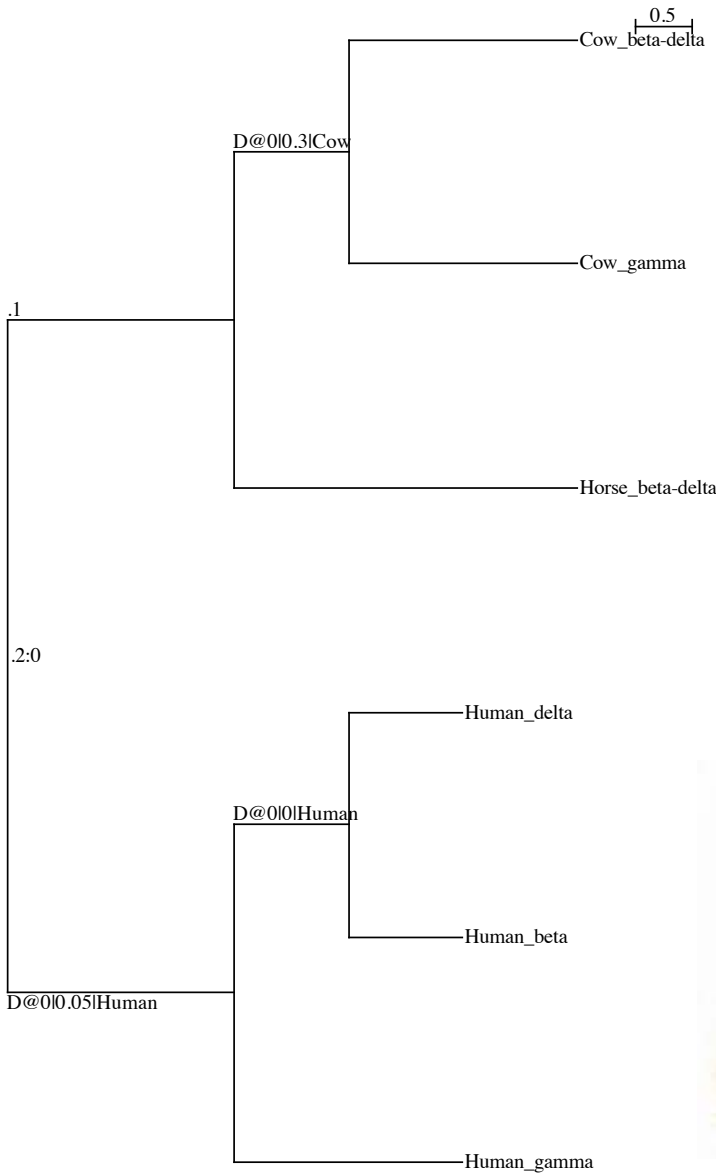


Hemoglobin/

```
((Human_gamma:2,(Human_beta:1,Human_delta:1)D@0|0.01|Human:1)D@0|0.05|Human:2,(Horse_beta-delta:3,(Cow_gamma:2,Cow_beta-delta:2)D@0|0.3|Cow:1).1:2).2:0;
```

ALEml HuHoCo.tree Hemoglobin.tree.ale tau=0 sample=10

(...).2:0;

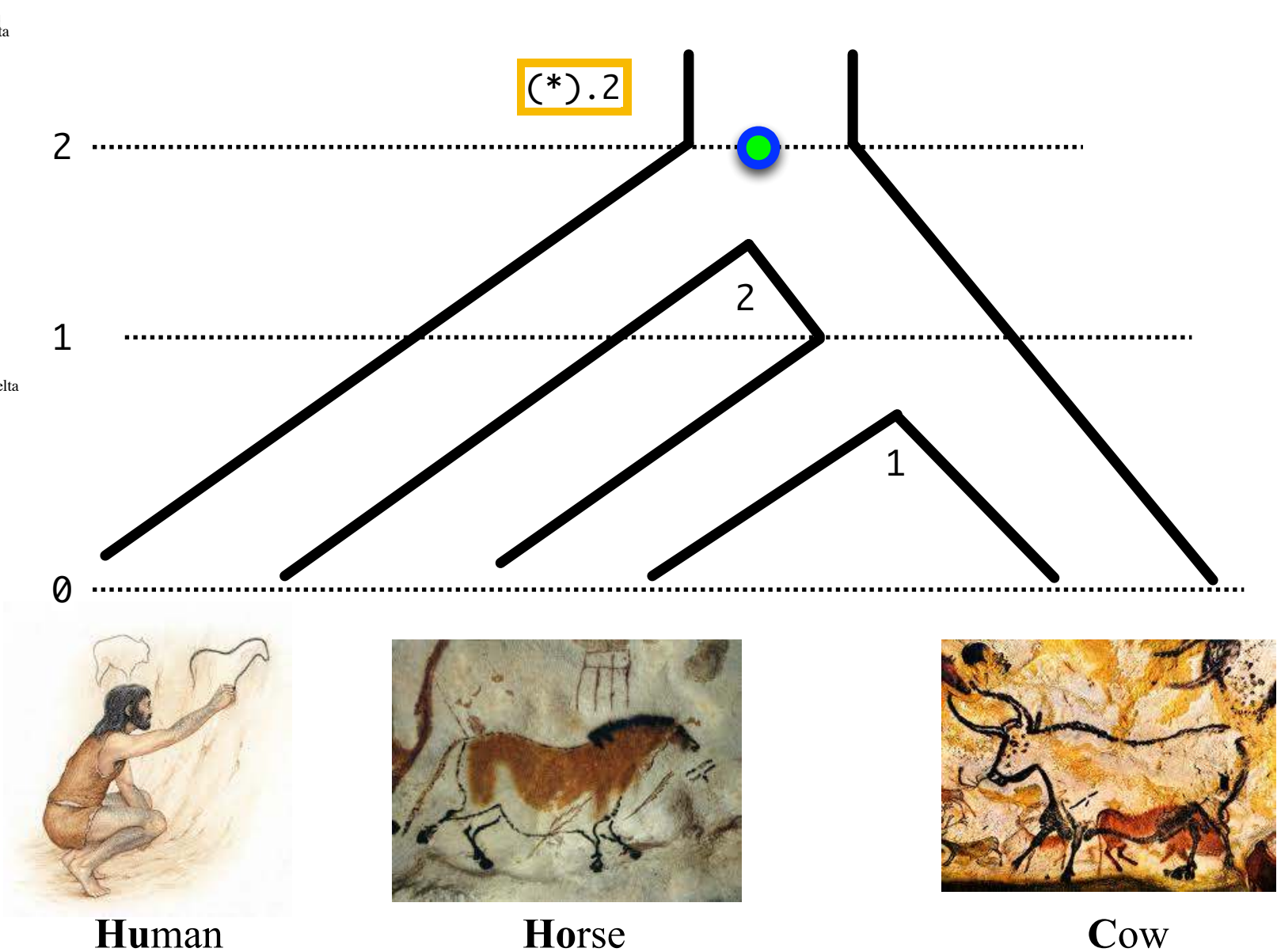
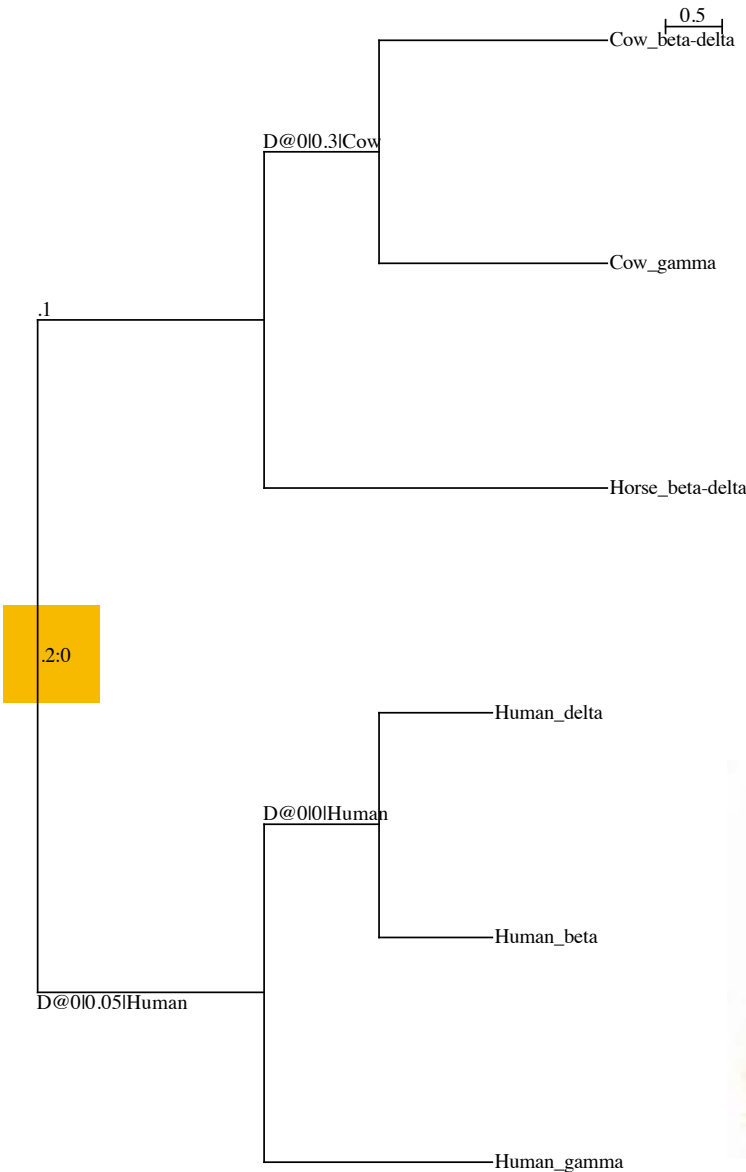


Hemoglobin/

```
((Human_gamma:2,(Human_beta:1,Human_delta:1)D@0|0.01|Human:1)D@0|0.05|Human:2,(Horse_beta-delta:3,(Cow_gamma:2,Cow_beta-delta:2)D@0|0.3|Cow:1).1:2).2:0;
```

ALEml HuHoCo.tree Hemoglobin.tree.ale tau=0 sample=10

(...).2:0;

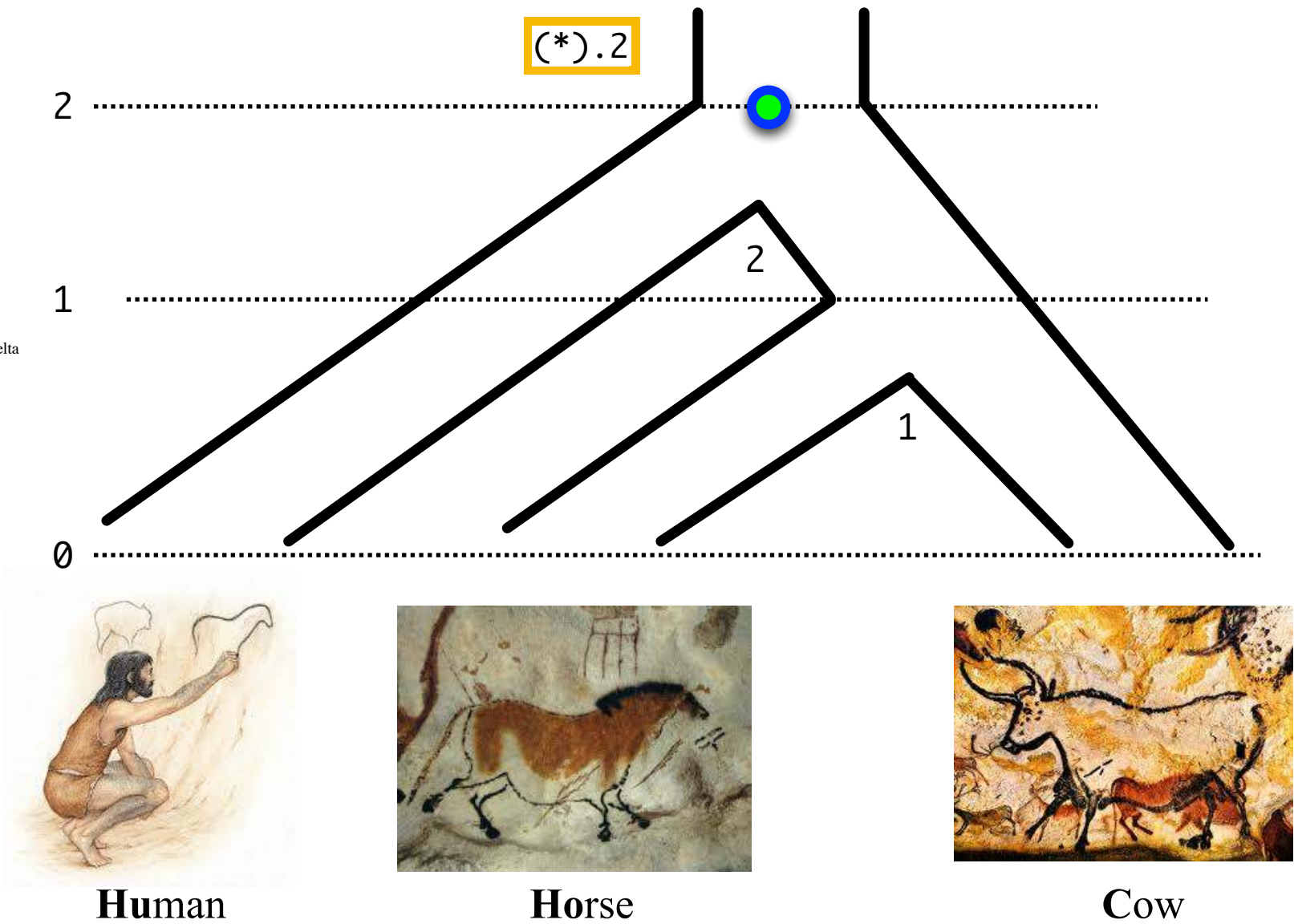
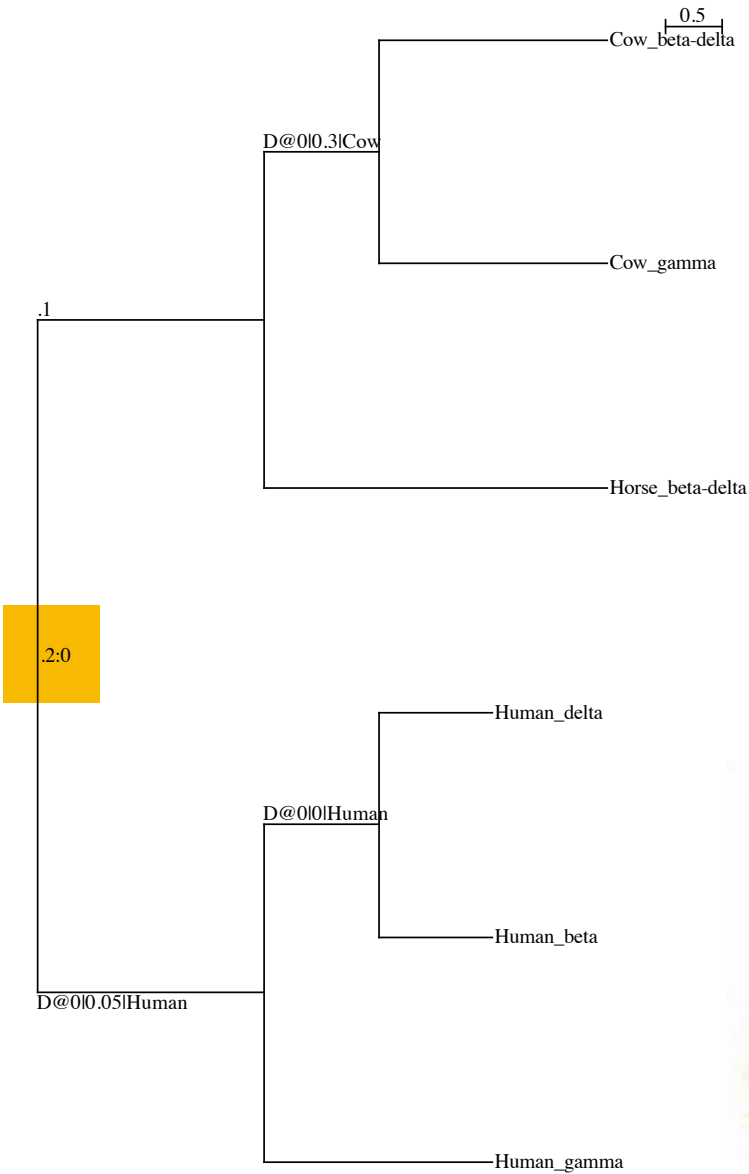


Hemoglobin/

((Human_gamma:2,(Human_beta:1,Human_delta:1)D@0|0.01|Human:1)D@0|0.05|Human:2,(Horse_beta-delta:3,(Cow_gamma:2,Cow_beta-delta:2)D@0|0.3|Cow:1).1:2).2:0;

ALEml HuHoCo.tree Hemoglobin.tree.ale tau=0 sample=10

(...).2:0; ((...)D@0|0.05|Human:2 , (...).1:2).2:0;).2:0;



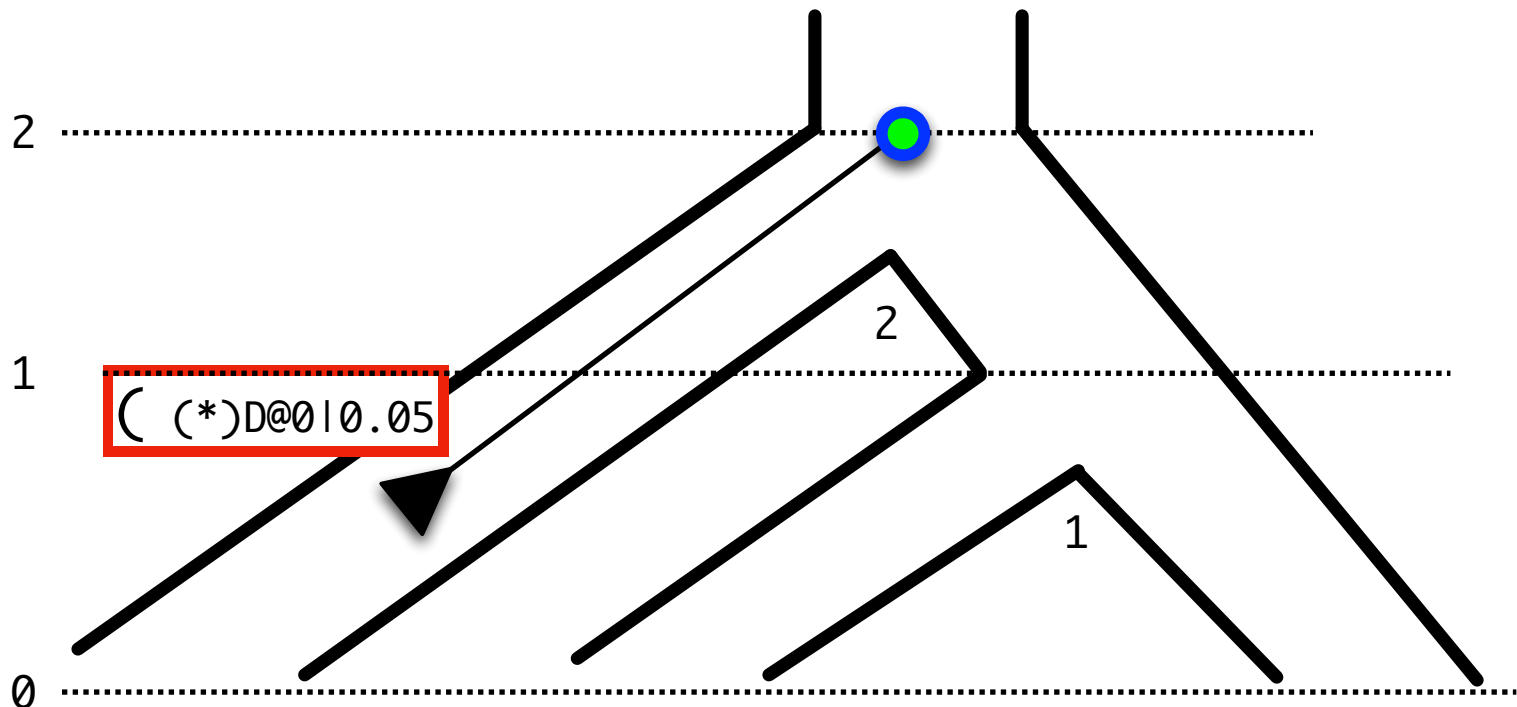
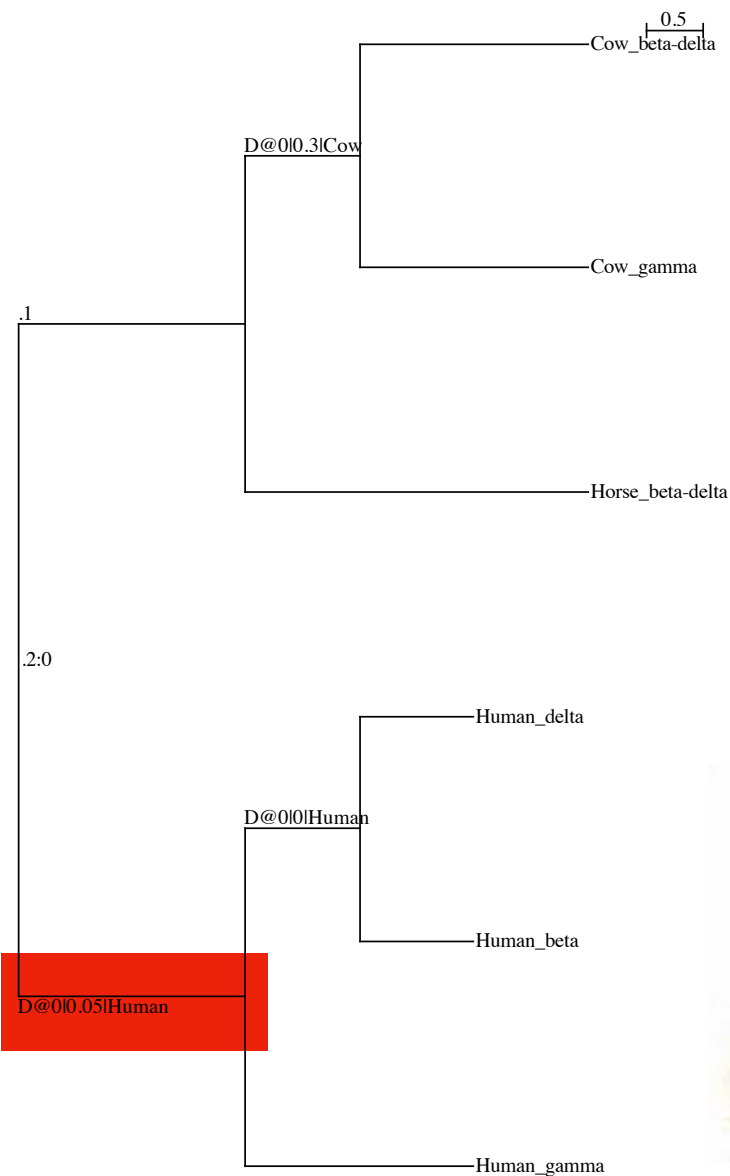
Hemoglobin/

((Human_gamma:2,(Human_beta:1,Human_delta:1)D@0|0.01|Human:1)D@0|0.05|Human:2,(Horse_beta-delta:3,(Cow_gamma:2,Cow_beta-delta:2)D@0|0.3|Cow:1).1:2).2:0;

ALEml HuHoCo.tree Hemoglobin.tree.ale tau=0 sample=10

(...).2:0;

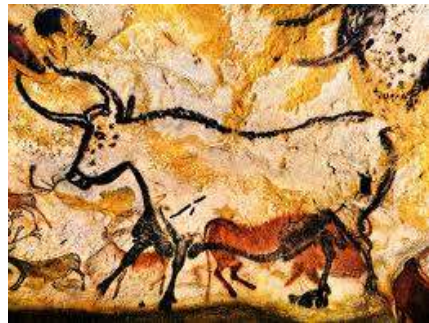
((...)D@0|0.05|Human:2 , (...).1:2).2:0;).2:0;



Human



Horse



Cow

Hemoglobin/

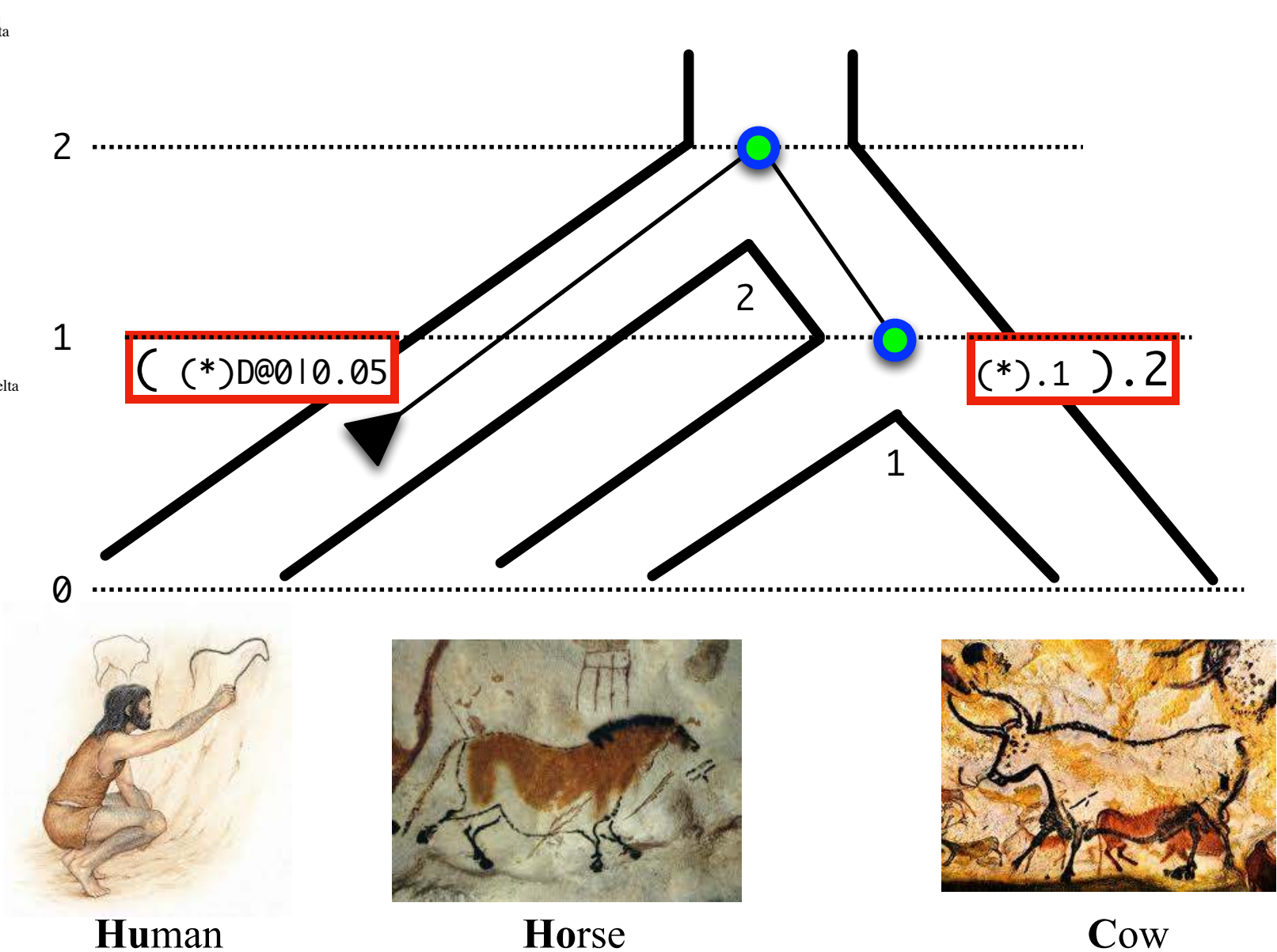
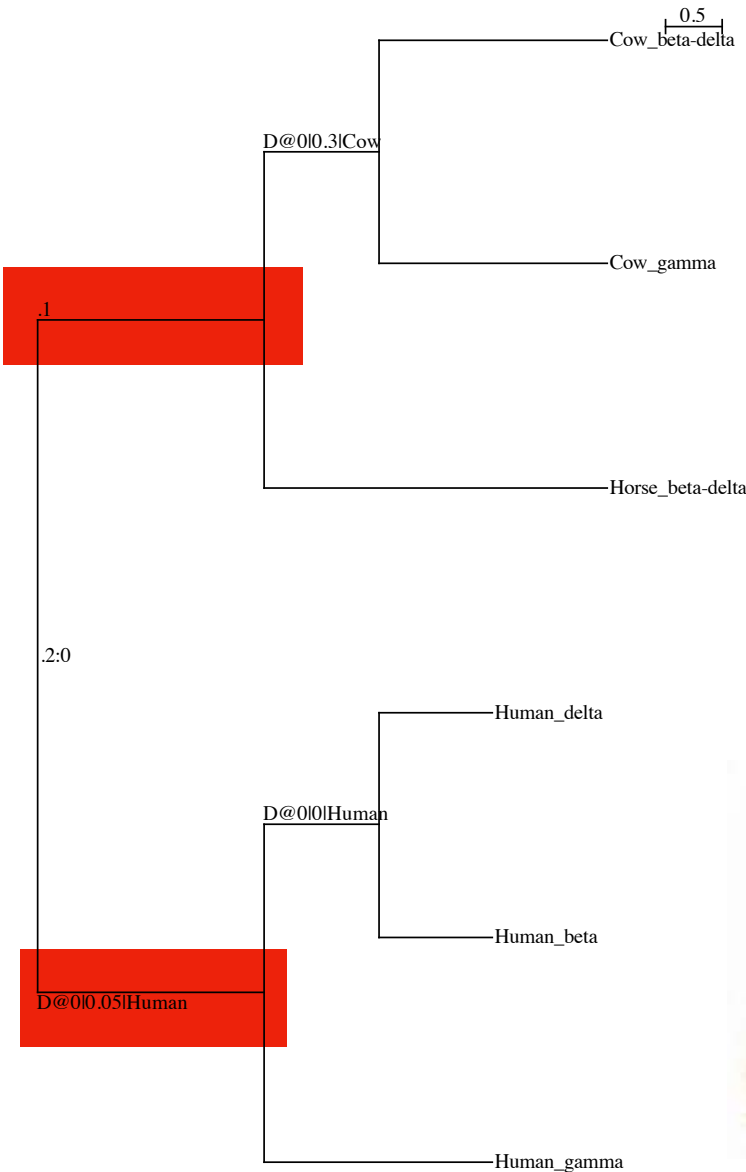
((Human_gamma:2,(Human_beta:1,Human_delta:1)D@0|0.01|Human:1)D@0|0.05|Human:2,(Horse_beta-delta:3,(Cow_gamma:2,Cow_beta-delta:2)D@0|0.3|Cow:1).1:2).2:0;

ALEml HuHoCo.tree Hemoglobin.tree.ale tau=0 sample=10

(...).2:0;

((...)D@0|0.05|Human:2

, (...).1:2).2:0;).2:0;



Hemoglobin/

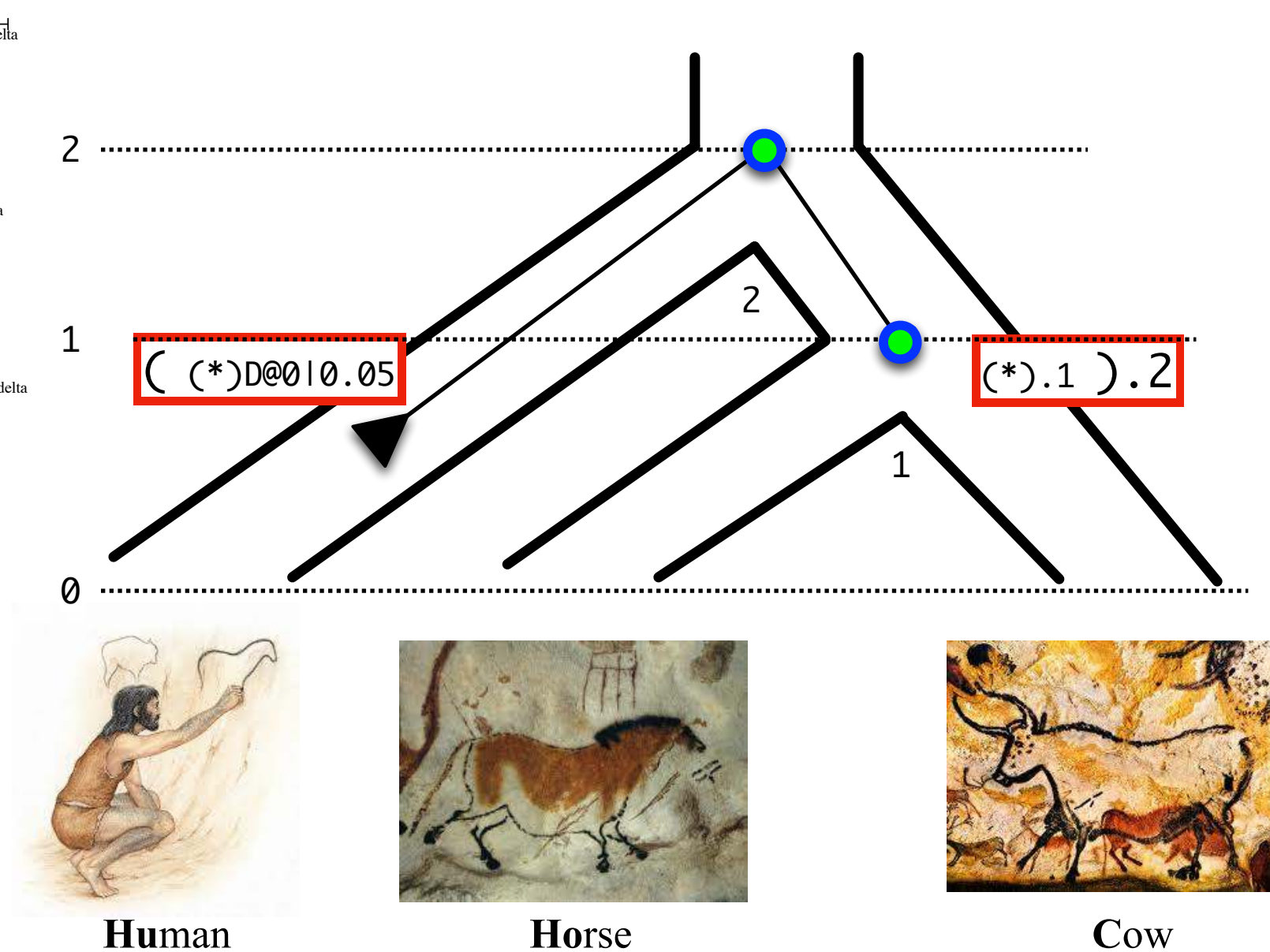
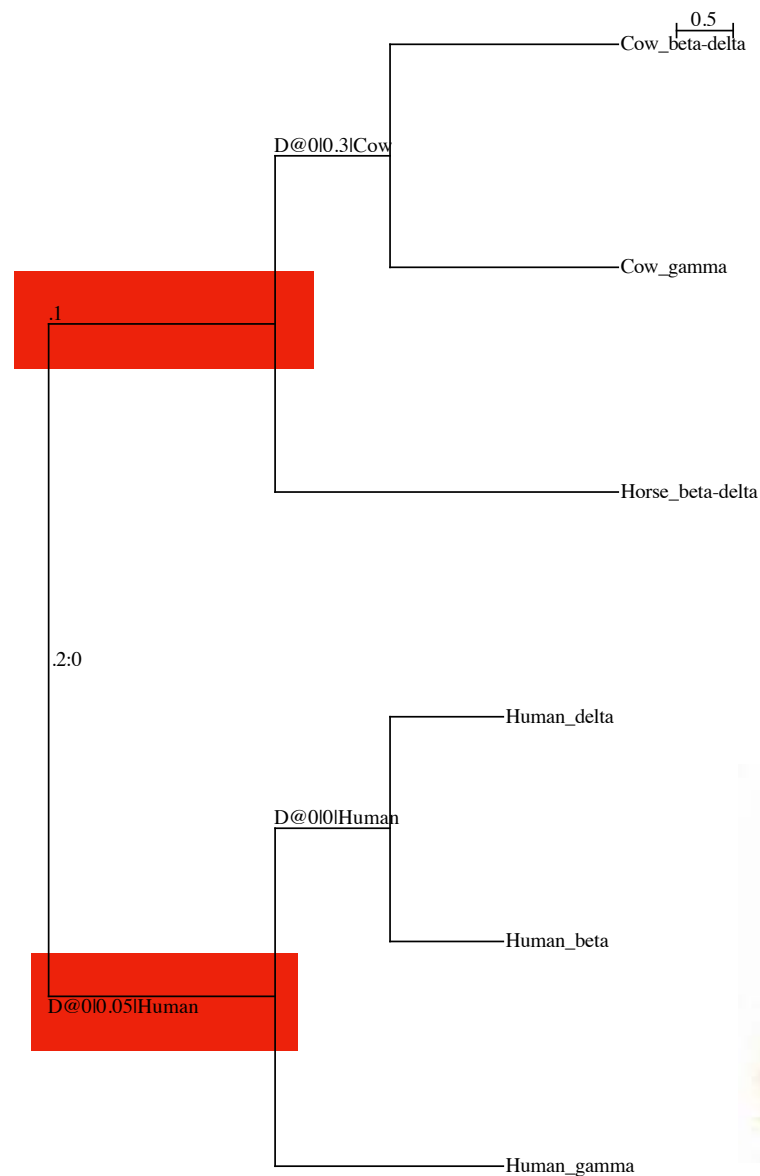
(C(Human_gamma:2,(HHuman_beta:1,Human_delta:1)D@0|0.01|Human:1)D@0|0.05|Human:2,(Horse_beta-delta:3,(Cow_gamma:2,Cow_beta-delta:2)D@0|0.3|Cow:1).1:2).2:0;

ALEmI HuHoCo.tree Hemoglobin.tree.ale tau=0 sample=10

```
(..).2:0;
```

$$((\text{Human_gamma}:2, (\text{Human_beta}:1, \text{Human_delta}:1)D@0|0.01|\text{Human}:1)D@0|0.05|\text{Human}:2),$$

```
, (Horse_beta-delta:3,(Cow_gamma:2,Cow_beta-delta:2)D@0|0.3|Cow:1).1:2 ).2:0; ).2:0;
```



Hemoglobin/

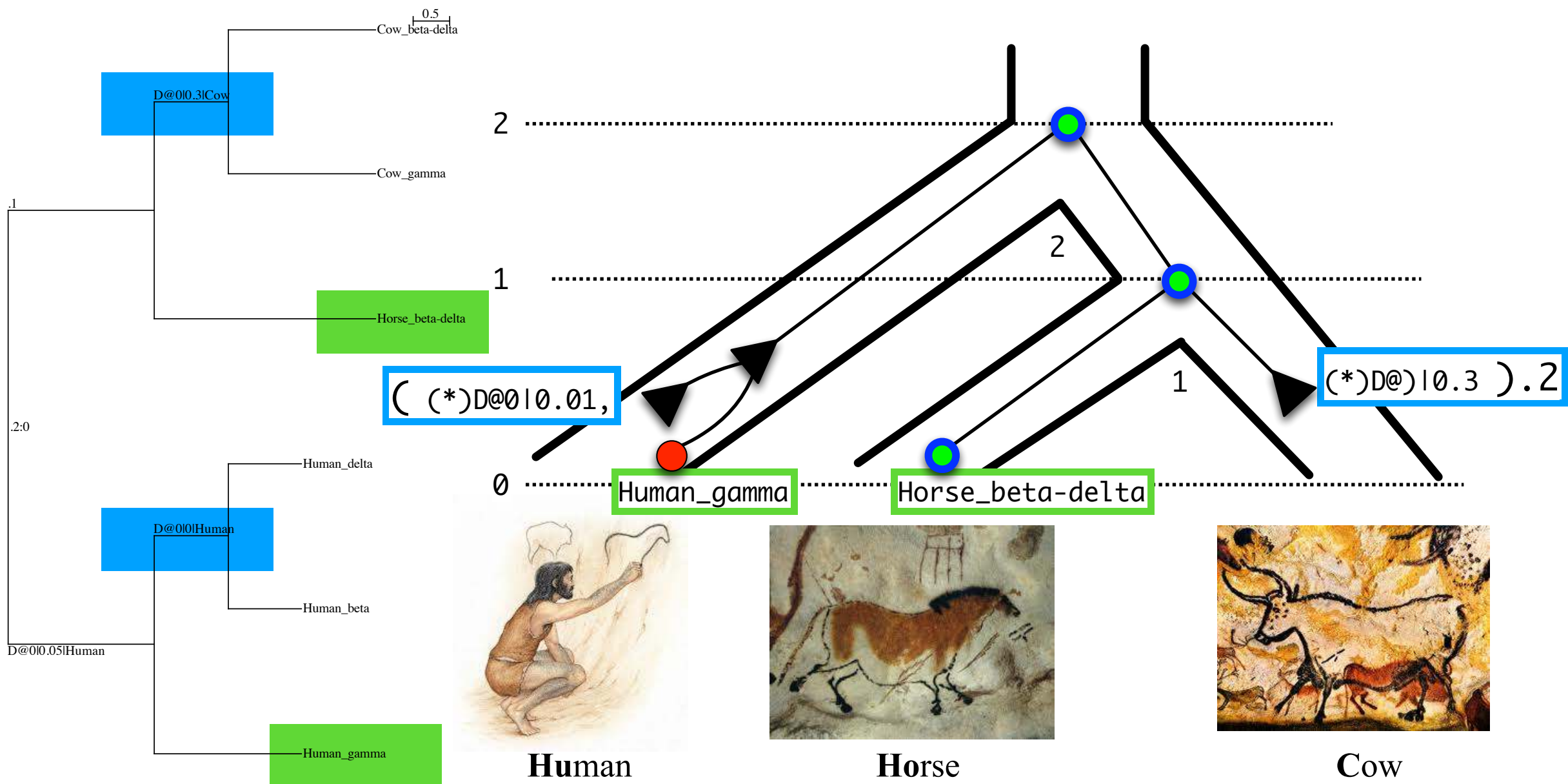
((Human_gamma:2,(Human_beta:1,Human_delta:1)D@0|0.01|Human:1)D@0|0.05|Human:2,(Horse_beta-delta:3,(Cow_gamma:2,Cow_beta-delta:2)D@0|0.3|Cow:1).1:2).2:0;

ALEml HuHoCo.tree Hemoglobin.tree.ale tau=0 sample=10

(...).2:0;

((Human_gamma:2,(Human_beta:1,Human_delta:1)D@0|0.01|Human:1)D@0|0.05|Human:2 ,

, (Horse_beta-delta:3,(Cow_gamma:2,Cow_beta-delta:2)D@0|0.3|Cow:1).1:2).2:0;).2:0;



Hemoglobin/

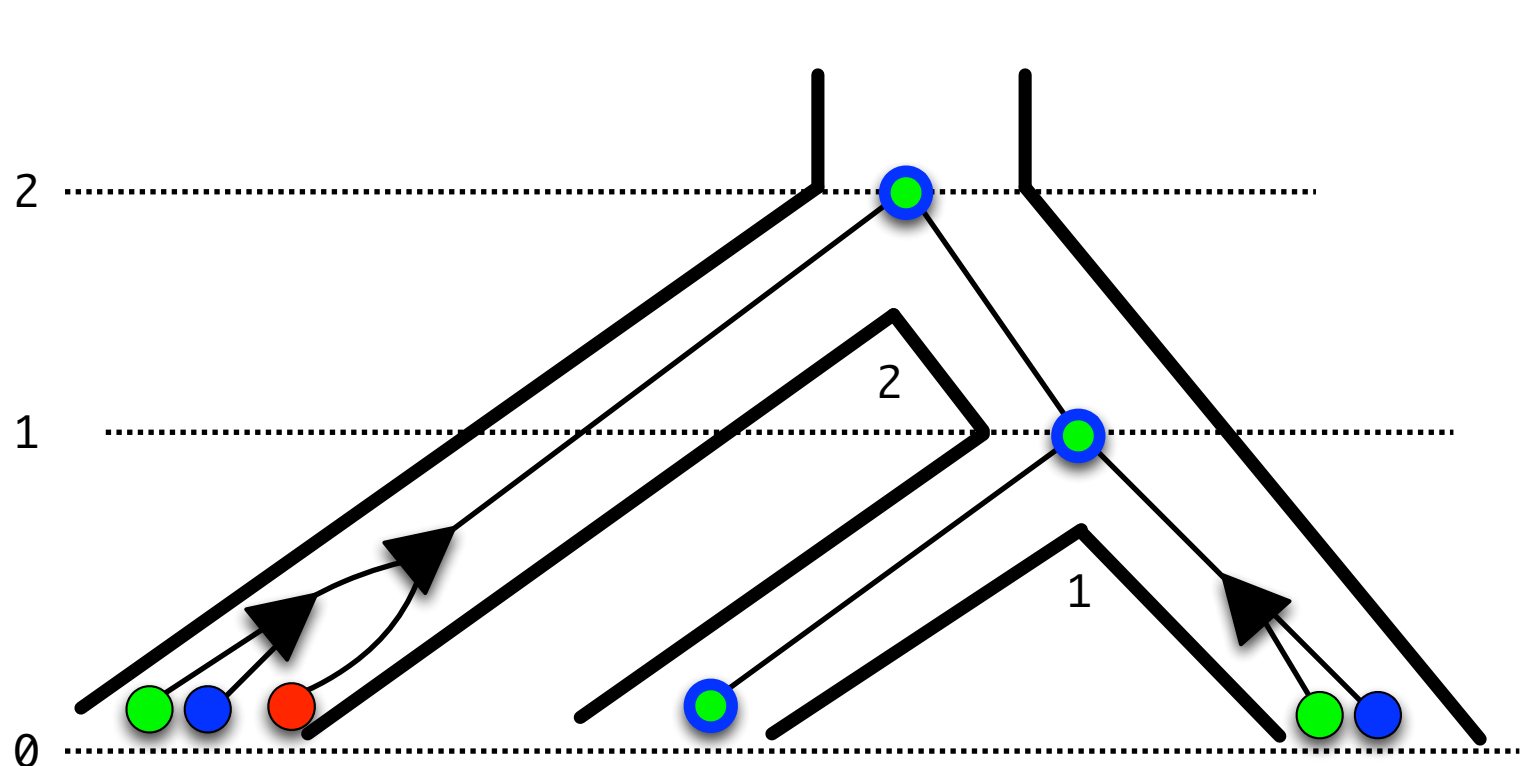
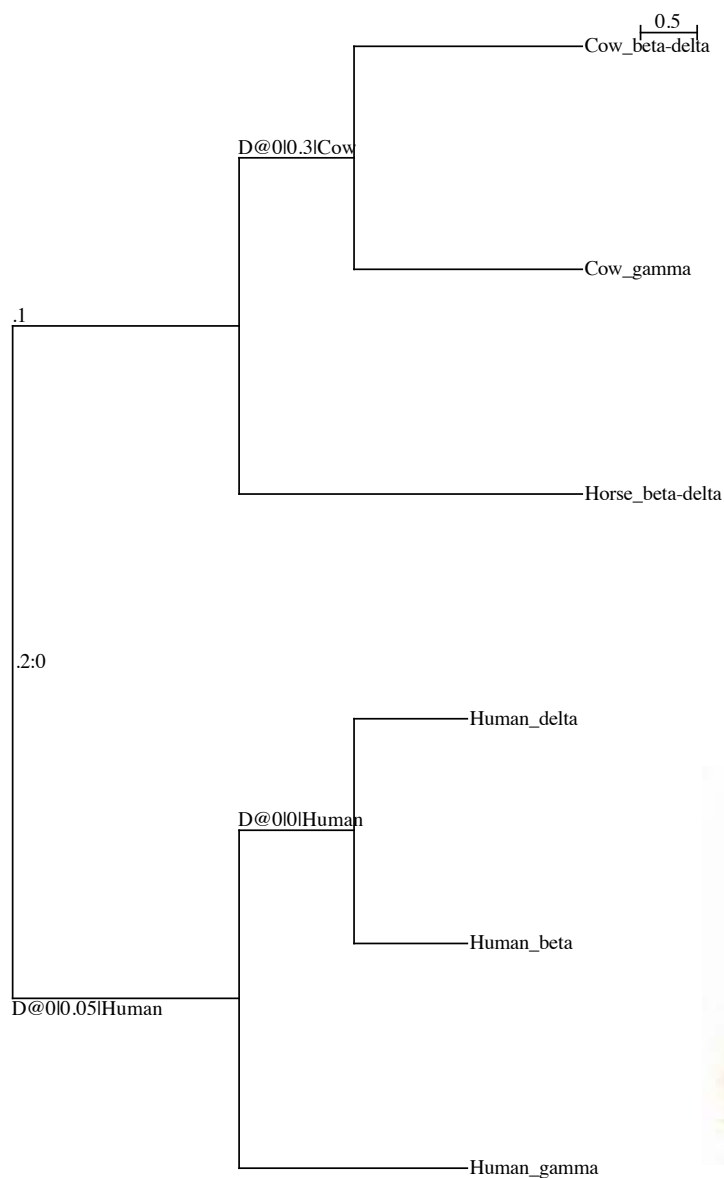
((Human_gamma:2,(Human_beta:1,Human_delta:1)D@0|0.01|Human:1)D@0|0.05|Human:2,(Horse_beta-delta:3,(Cow_gamma:2,Cow_beta-delta:2)D@0|0.3|Cow:1).1:2).2:0;

ALEml HuHoCo.tree Hemoglobin.tree.ale tau=0 sample=10

(...).2:0;

((Human_gamma:2,(Human_beta:1,Human_delta:1)D@0|0.01|Human:1)D@0|0.05|Human:2 ,

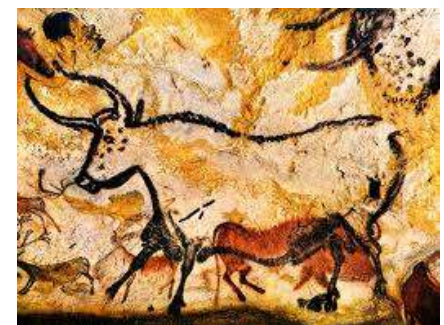
, (Horse_beta-delta:3,(Cow_gamma:2,Cow_beta-delta:2)D@0|0.3|Cow:1).1:2).2:0;).2:0;



Human



Horse



Cow

Hemoglobin/

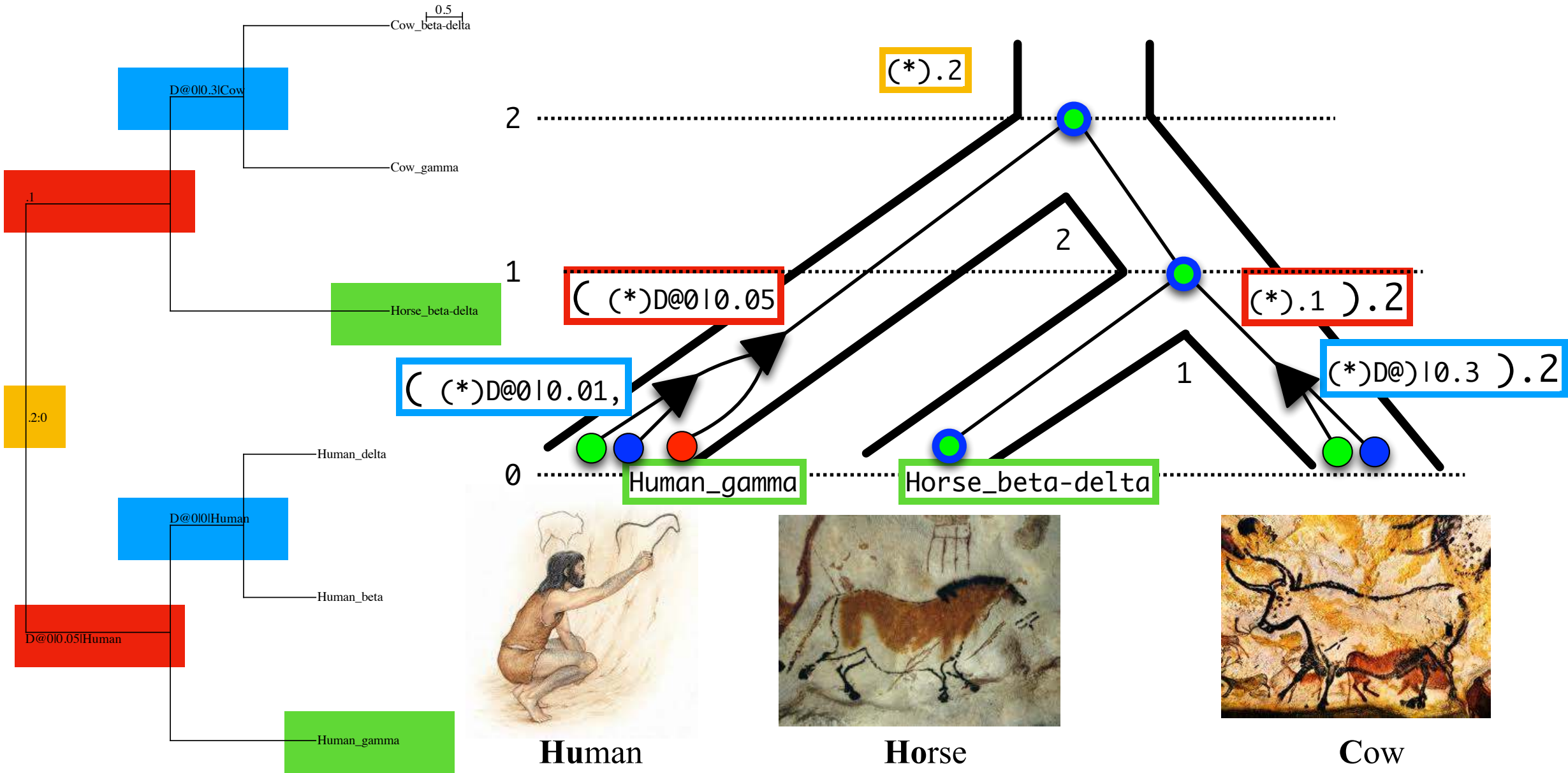
ALEml HuHoCo.tree Hemoglobin.tree.ale tau=0 sample=10

((Human_gamma:2,(Human_beta:1,Human_delta:1)D@0|0.01|Human:1)D@0|0.05|Human:2,(Horse_beta-delta:3,(Cow_gamma:2,Cow_beta-delta:2)D@0|0.3|Cow:1).1:2).2:0;

# of	Duplications	Transfers	Losses	Speciations
Total	3	0	0	2

# of	Duplications	Transfers	Losses	copies	
S_terminal_branch	Cow	1	0	0	2
S_terminal_branch	Horse	0	0	0	1
S_terminal_branch	Human	2	0	0	3
S_internal_branch	1	0	0	0	1
S_internal_branch	2	0	0	0	1

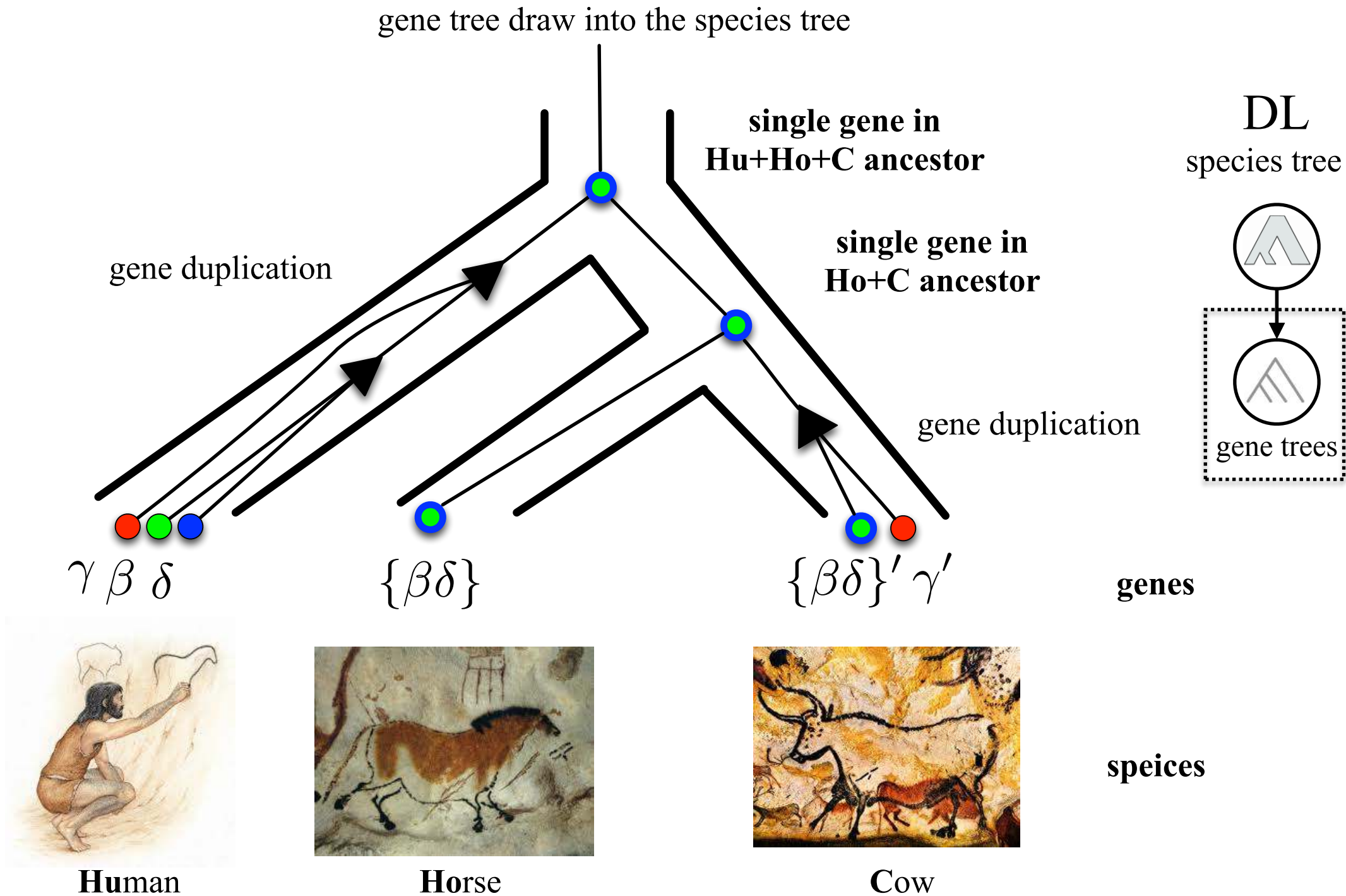
← mean number of events per branch in sampled reconciliations



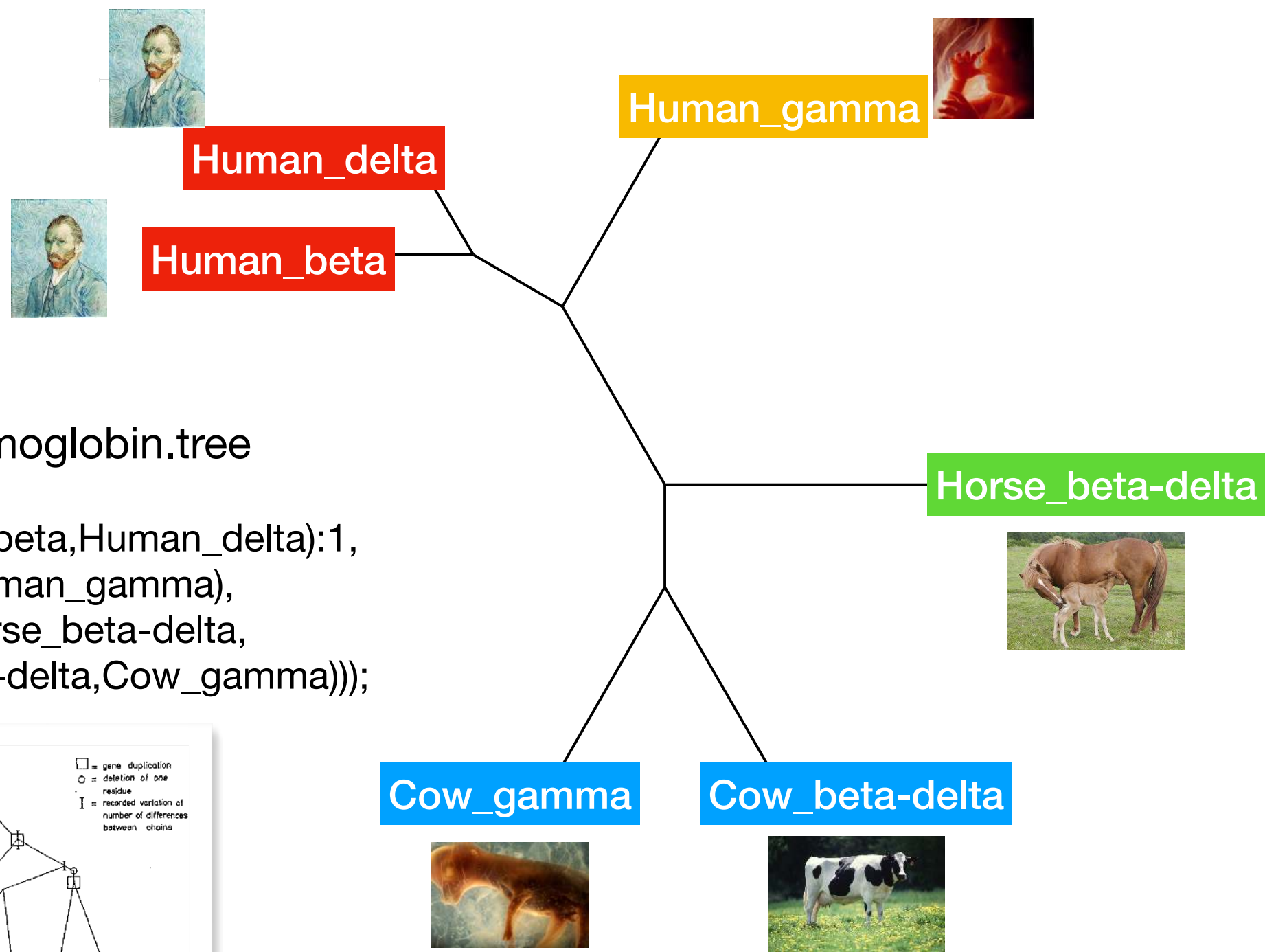
view the tree using either
phylo.io, seaview or FigTree!

The stories gene families can be complicated

The story of each gene family consist of a unique series of evolutionary events that often results in a change of copy number and shifts in function.

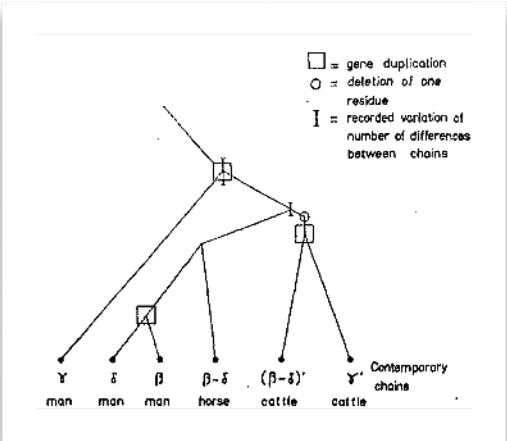


Hemoglobin gene tree

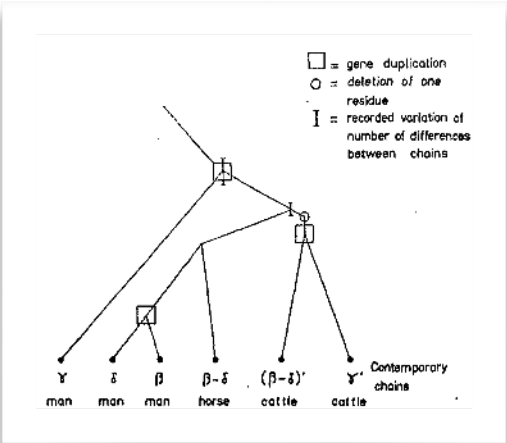
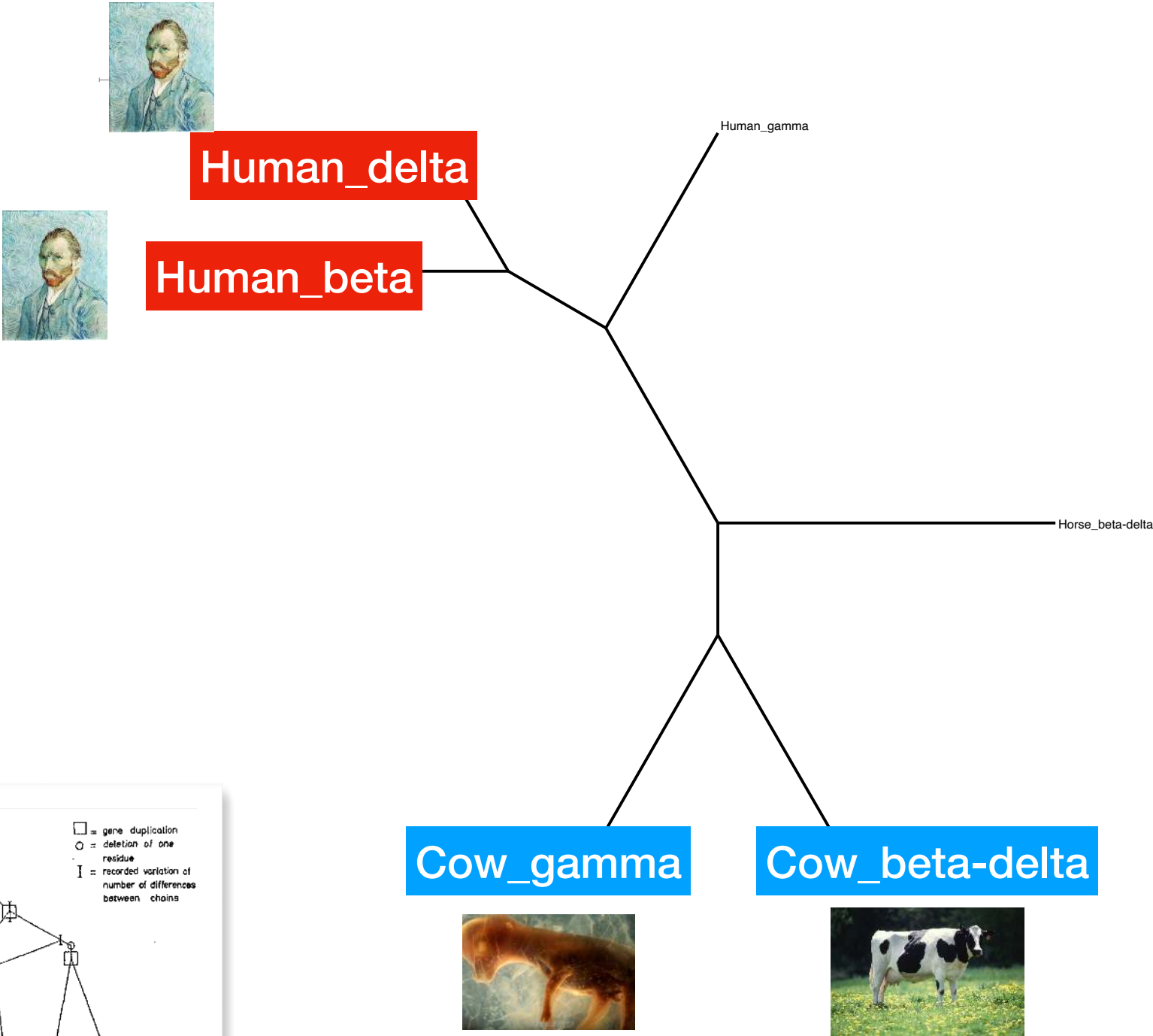


Hemoglobin.tree

(((Human_beta,Human_delta):1,
Human_gamma),
(Horse_beta-delta,
(Cow_beta-delta,Cow_gamma))));

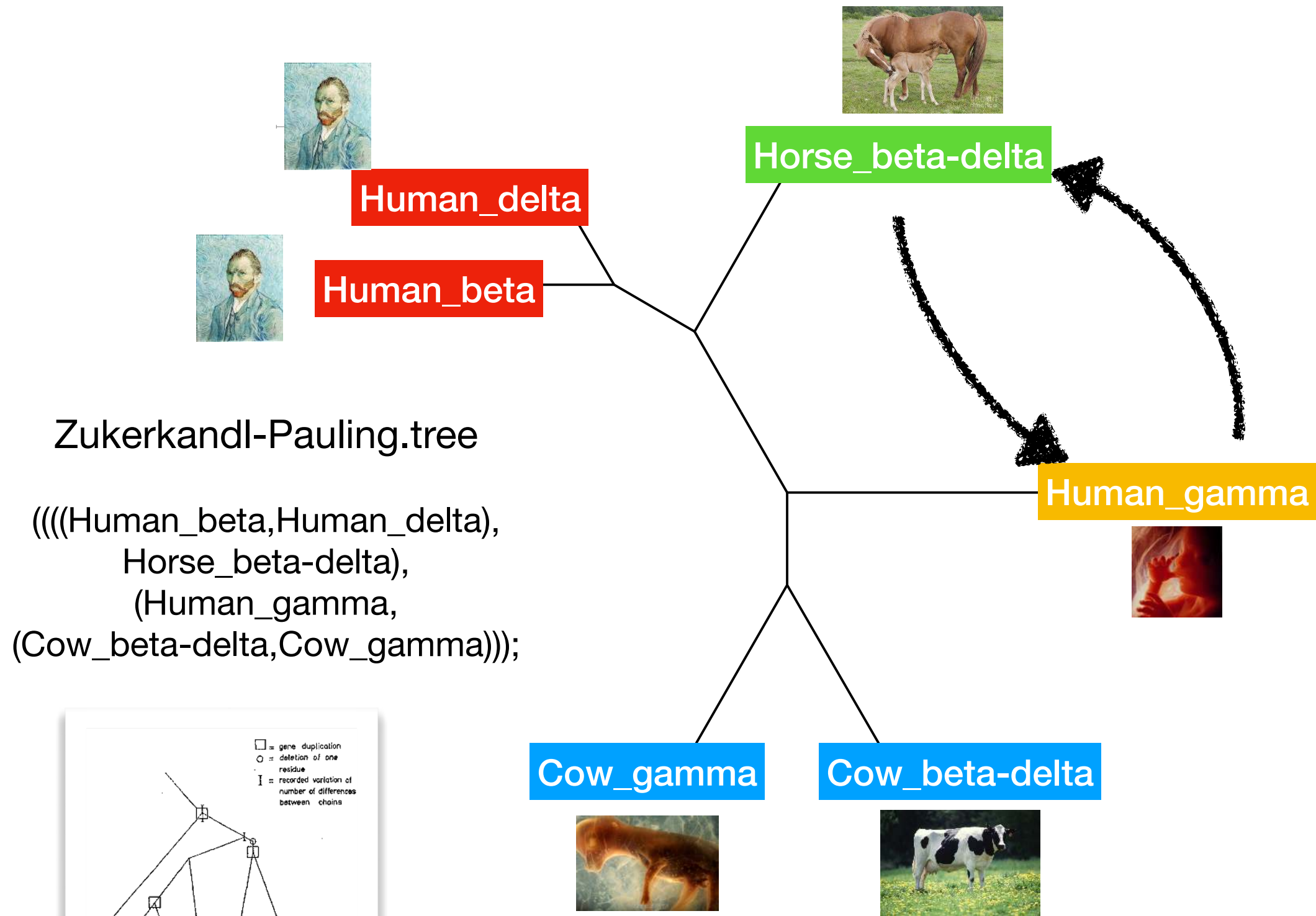


Hemoglobin/



Zukerkandl & Pauling 1965

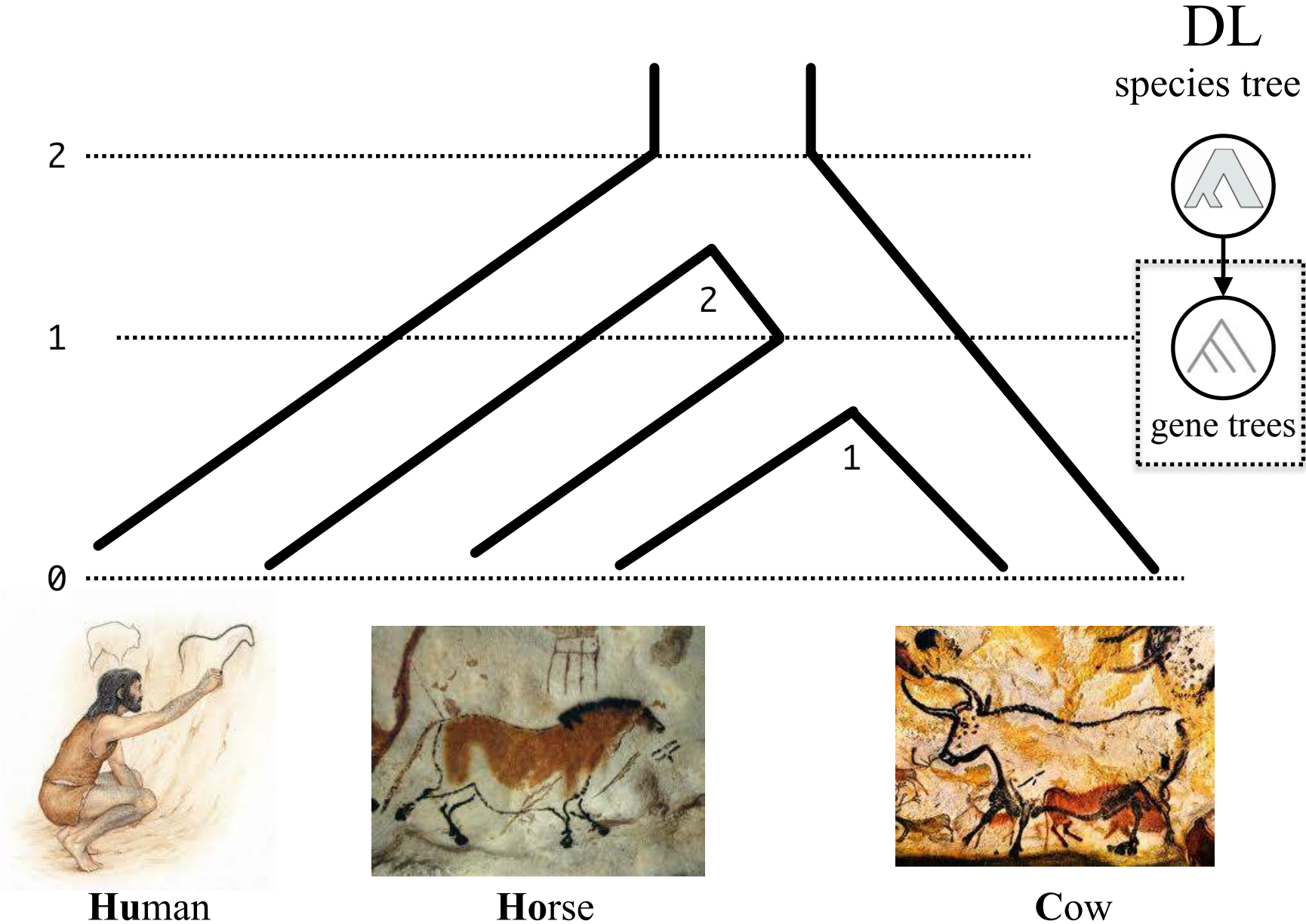
The first ever gene tree



Hemoglobin/

ALEobserve Zukerkandl-Pauling.tree

ALEml HuHoCo.tree **Zukerkandl-Pauling.tree.ale** tau=0 sample=10



Hemoglobin/

ALEml HuHoCo.tree Zukerkandl-Pauling.tree.ale tau=0 sample=10

#ALEml using ALE v0.5 by Szollosi GJ et al.; ssolo@elte.hu; CC BY-SA 3.0;

S: (Human:1,(Horse:0.5,Cow:0.5)1:0.5)2; ← species tree

Input ale from: HuHoCo.tree_Zukerkandl-Pauling.tree.ale
>logl: -10.0385 ← ML log-likelihood (tau=0)

rate of Duplications Transfers Losses
ML 0.721451 0 0.699903 ← ML rates (tau=0)

10 reconciled G-s:

((Human_gamma:5,(Cow_beta-delta:1,Cow_gamma:1).1D@0|0.2|Cow:2).2:1,(Horse_beta-delta:1:2,(Human_beta:1,Human_delta:1)D@1|0.55|Human:1).2:1)D@2|1.05|2:0;
(Horse_beta-delta:2.1:2,((Human_beta:1,Human_delta:1).2D@1|0.55|Human:1,(Human_gamma:5,(Cow_beta-delta:1,Cow_gamma:1).1D@0|0.1|Cow:2).2:1)D@2|1|2:2)D@2|1.35|2:0;
((Human_gamma:5,(Cow_beta-delta:1,Cow_gamma:1).1D@0|0.45|Cow:2).2:1,(Horse_beta-delta:1:2,(Human_beta:1,Human_delta:1)D@0|0.05|Human:1).2:1)D@2|1.2|2:0;
((Human_gamma:5,(Cow_beta-delta:1,Cow_gamma:1).1D@0|0.4|Cow:2).2:1,(Horse_beta-delta:1:2,(Human_beta:1,Human_delta:1)D@0|0.15|Human:1).2:1)D@2|1.45|2:0;
((Human_gamma:5,(Cow_beta-delta:1,Cow_gamma:1).1D@0|0.3|Cow:2).2:1,(Horse_beta-delta:1:2,(Human_beta:1,Human_delta:1)D@0|0|Human:1).2:1)D@2|1|2:0;
(Horse_beta-delta:2.1:2,((Human_beta:1,Human_delta:1).2D@0|0.4|Human:1,(Human_gamma:5,(Cow_beta-delta:1,Cow_gamma:1).1D@0|0.05|Cow:2).2:1)D@2|1|2:2)D@2|1.3|2:0;
((Human_gamma:5,(Cow_beta-delta:1,Cow_gamma:1).1D@0|0.1|Cow:2).2:1,(Horse_beta-delta:1:2,(Human_beta:1,Human_delta:1)D@0|0.1|Human:1).2:1)D@2|1.35|2:0;
((Human_gamma:5,(Cow_beta-delta:1,Cow_gamma:1).1D@0|0.15|Cow:2).2:1,(Horse_beta-delta:1:2,(Human_beta:1,Human_delta:1)D@0|0.05|Human:1).2:1)D@2|1.1|2:0;
((Human_gamma:5,(Cow_beta-delta:1:1,Cow_gamma:1:1)D@1|0.5|1:2).2:1,(Horse_beta-delta:1:2,(Human_beta:1,Human_delta:1)D@0|0.2|Human:1).2:1)D@2|1.05|2:0;
((Human_gamma:5,(Cow_beta-delta:1,Cow_gamma:1).1D@0|0.15|Cow:2).2:1,(Horse_beta-delta:1:2,(Human_beta:1,Human_delta:1)D@1|0.55|Human:1).2:1)D@2|1.05|2:0;

# of	Duplications	Transfers	Losses	Speciations
Total	3.2	0	2.5	4.3

# of	Duplications	Transfers	Losses	copies	
S_terminal_branch	Cow	0.9	0	1	2
S_terminal_branch	Horse	0	0	1.1	1
S_terminal_branch	Human	1	0	0.2	3
S_internal_branch	1	0.1	0	0.2	2.1
S_internal_branch	2	1.2	0	0	2.2

HuHoCo.tree_Zukerkandl-Pauling.tree.ale.ml_rec (END)

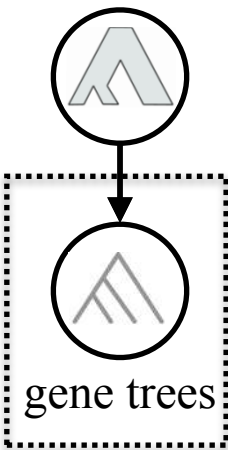
mean number of events
per branch in sampled
reconciliations

10 random reconciled gene trees
sampled according to joint likelihood

$$\sum_{\text{gene trees}} P(\text{events} | \text{gene tree}) P(\text{gene tree} | \text{species tree})$$

sum over reconciled
gene trees

DL
species tree



Hemoglobin/

ALEml HuHoCo.tree **Zukerkandl-Pauling.tree.ale** tau=0 sample=10

```
#ALEml using ALE v0.5 by Szollosi GJ et al.; ssolo@elte.hu; CC BY-SA 3.0;
S:      (Human:1,(Horse:0.5,Cow:0.5)1:0.5)2;
```

```
Input ale from: Zukerkandl-Pauling.tree.ale
>logl: -10.0385
rate of Duplications  Transfers      Losses
ML      0.721451      0             0.699903

..
# of      Duplications  Transfers      Losses  Speciations
Total    3.4           0             2.9    4.5

# of      Duplications  Transfers      Losses  copies
S_terminal_branch      Cow    0.9      0      1      2
S_terminal_branch      Horse  0        0      1.1    1
S_terminal_branch      Human  1        0      0.4    3
S_internal_branch      1      0.1      0      0.4    2.1
S_internal_branch      2      1.4      0      0      2.4
HuHoCo.tree_Zukerkandl-Pauling.tree.ale.ml_rec (END)
```

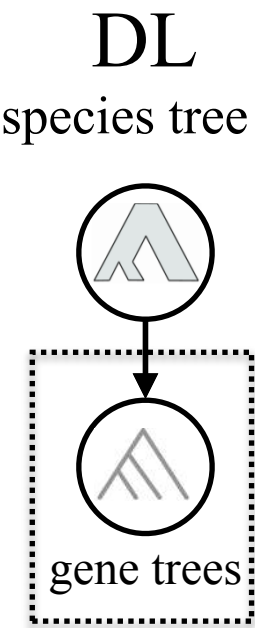
ALEml HuHoCo.tree **Hemoglobin.tree.ale** tau=0 sample=10

```
#ALEml using ALE v0.5 by Szollosi GJ et al.; ssolo@elte.hu; CC BY-SA 3.0;
S:      (Human:1,(Horse:0.5,Cow:0.5)1:0.5)2;
```

```
Input ale from: Hemoglobin.tree.ale
>logl: -4.45543
rate of Duplications  Transfers      Losses
ML      0.76128      0             1e-10

..
# of      Duplications  Transfers      Losses  Speciations
Total    3             0             0      2

# of      Duplications  Transfers      Losses  copies
S_terminal_branch      Cow    1        0      0      2
S_terminal_branch      Horse  0        0      0      1
S_terminal_branch      Human  2        0      0      3
S_internal_branch      1      0        0      0      1
S_internal_branch      2      0        0      0      1
HuHoCo.tree_Hemoglobin.tree.ale.ml_rec (END)
```



Hemoglobin/

ALEml HuHoCo.tree Zukerkandl-Pauling.tree.ale **delta=0.01 lambda=0.01** tau=0 sample=10

#ALEml using ALE v0.5 by Szollosi GJ et al.; ssolo@elte.hu; CC BY-SA 3.0;

S: (Human:1,(Horse:0.5,Cow:0.5)1:0.5)2;

Input ale from: Zukerkandl-Pauling.tree.ale

>logl: -25.1341

rate of	Duplications	Transfers	Losses
ML	0.01	0	0.01

..				
# of	Duplications	Transfers	Losses	Speciations
Total	3	0	2	4

# of	Duplications	Transfers	Losses	copies	
S_terminal_branch	Cow	1	0	1	2
S_terminal_branch	Horse	0	0	1	1
S_terminal_branch	Human	1	0	0	3
S_internal_branch	1	0	0	0	2
S_internal_branch	2	1	0	0	2

HuHoCo.tree_Zukerkandl-Pauling.tree.ale.ml_rec (END)

ALEml HuHoCo.tree Zukerkandl-Pauling.tree.ale tau=0 sample=10

Input ale from: Zukerkandl-Pauling.tree.ale

>logl: -10.0385

rate of	Duplications	Transfers	Losses
ML	0.721451	0	0.699903

..				
# of	Duplications	Transfers	Losses	Speciations
Total	3.4	0	2.9	4.5

# of	Duplications	Transfers	Losses	copies	
S_terminal_branch	Cow	0.9	0	1	2
S_terminal_branch	Horse	0	0	1.1	1
S_terminal_branch	Human	1	0	0.4	3
S_internal_branch	1	0.1	0	0.4	2.1
S_internal_branch	2	1.4	0	0	2.4

HuHoCo.tree_Zukerkandl-Pauling.tree.ale.ml_rec (END)

ALEml HuHoCo.tree Hemoglobin.tree.ale tau=0 sample=10

#ALEml using ALE v0.5 by Szollosi GJ et al.; ssolo@elte.hu; CC BY-SA 3.0;

S: (Human:1,(Horse:0.5,Cow:0.5)1:0.5)2;

Input ale from: Hemoglobin.tree.ale

>logl: -4.45543

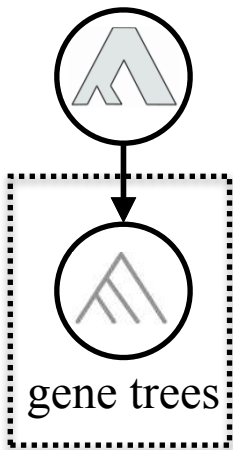
rate of	Duplications	Transfers	Losses
ML	0.76128	0	1e-10

..				
# of	Duplications	Transfers	Losses	Speciations
Total	3	0	0	2

# of	Duplications	Transfers	Losses	copies	
S_terminal_branch	Cow	1	0	0	2
S_terminal_branch	Horse	0	0	0	1
S_terminal_branch	Human	2	0	0	3
S_internal_branch	1	0	0	0	1
S_internal_branch	2	0	0	0	1

HuHoCo.tree_Hemoglobin.tree.ale.ml_rec (END)

DL
species tree



Hemoglobin/

ALEml HuHoCo.tree Zukerkandl-Pauling.tree.ale **delta=0.01 lambda=0.01 tau=0 sample=10**

#ALEml using ALE v0.5 by Szollosi GJ et al.; ssolo@elte.hu; CC BY-SA 3.0;

S: (Human:1,(Horse:0.5,Cow:0.5)1:0.5)2;

Input ale from: Zukerkandl-Pauling.tree.ale

>logl: -25.1341

rate of	Duplications	Transfers	Losses
ML	0.01	0	0.01

10 reconciled G-s:

((Human_gamma:5,(Cow_beta-delta:1,Cow_gamma:1).1D@0|0|Cow:2).2:1,(Horse_beta-delta.1:2,(Human_beta:1,Human_delta:1)D@1|0.5|Human:1).2:1)D@2|1.45|2:0;
((Human_gamma:5,(Cow_beta-delta:1,Cow_gamma:1).1D@0|0.15|Cow:2).2:1,(Horse_beta-delta.1:2,(Human_beta:1,Human_delta:1)D@0|0.05|Human:1).2:1)D@2|1.3|2:0;
((Human_gamma:5,(Cow_beta-delta:1,Cow_gamma:1).1D@0|0.45|Cow:2).2:1,(Horse_beta-delta.1:2,(Human_beta:1,Human_delta:1)D@0|0.45|Human:1).2:1)D@2|1.35|2:0;
((Human_gamma:5,(Cow_beta-delta:1,Cow_gamma:1).1D@0|0.05|Cow:2).2:1,(Horse_beta-delta.1:2,(Human_beta:1,Human_delta:1)D@1|0.55|Human:1).2:1)D@2|1.65|2:0;
((Human_gamma:5,(Cow_beta-delta:1,Cow_gamma:1).1D@0|0.45|Cow:2).2:1,(Horse_beta-delta.1:2,(Human_beta:1,Human_delta:1)D@1|0.95|Human:1).2:1)D@2|1.55|2:0;
((Human_gamma:5,(Cow_beta-delta:1,Cow_gamma:1).1D@0|0.25|Cow:2).2:1,(Horse_beta-delta.1:2,(Human_beta:1,Human_delta:1)D@0|0.1|Human:1).2:1)D@2|1.05|2:0;
((Human_gamma:5,(Cow_beta-delta:1,Cow_gamma:1).1D@0|0|Cow:2).2:1,(Horse_beta-delta.1:2,(Human_beta:1,Human_delta:1)D@0|0.25|Human:1).2:1)D@2|1.6|2:0;
((Human_gamma:5,(Cow_beta-delta:1,Cow_gamma:1).1D@0|0.35|Cow:2).2:1,(Horse_beta-delta.1:2,(Human_beta:1,Human_delta:1)D@0|0.05|Human:1).2:1)D@2|1.35|2:0;
((Human_gamma:5,(Cow_beta-delta:1,Cow_gamma:1).1D@0|0.15|Cow:2).2:1,(Horse_beta-delta.1:2,(Human_beta:1,Human_delta:1)D@1|0.55|Human:1).2:1)D@2|1.1|2:0;
((Human_gamma:5,(Cow_beta-delta:1,Cow_gamma:1).1D@0|0.15|Cow:2).2:1,(Horse_beta-delta.1:2,(Human_beta:1,Human_delta:1)D@0|0.25|Human:1).2:1)D@2|1.7|2:0;

# of	Duplications	Transfers	Losses	Speciations
Total	3	0	2	4

# of	Duplications	Transfers	Losses	copies	
S_terminal_branch	Cow	1	0	1	2
S_terminal_branch	Horse	0	0	1	1
S_terminal_branch	Human	1	0	0	3
S_internal_branch	1	0	0	0	2
S_internal_branch	2	1	0	0	2

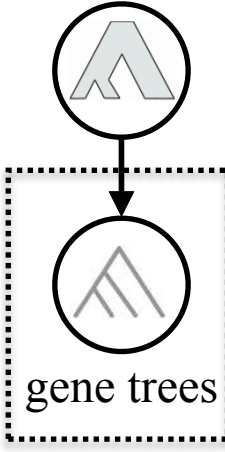
HuHoCo.tree_Zukerkandl-Pauling.tree.ale.ml_rec (END)

species tree

log-likelihood

fixed rates

DL
species tree



10 random reconciled gene trees
sampled according to joint likelihood

mean number of events
per branch in sampled
reconciliations

$$\sum_{\text{reconciled gene trees}} P(\text{events} | \text{gene tree}) P(\text{gene tree} | \text{species tree})$$

sum over reconciled
gene trees

Hemoglobin/

ALEml HuHoCo.tree Zukerkandl-Pauling.tree.ale **delta=0.01 lambda=0.01** tau=0 sample=10

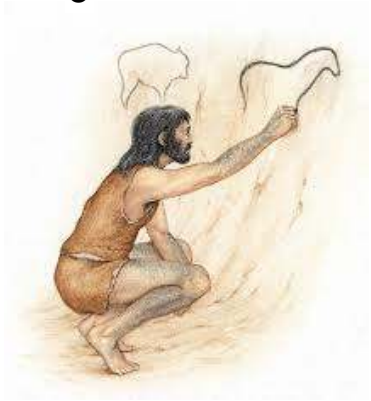
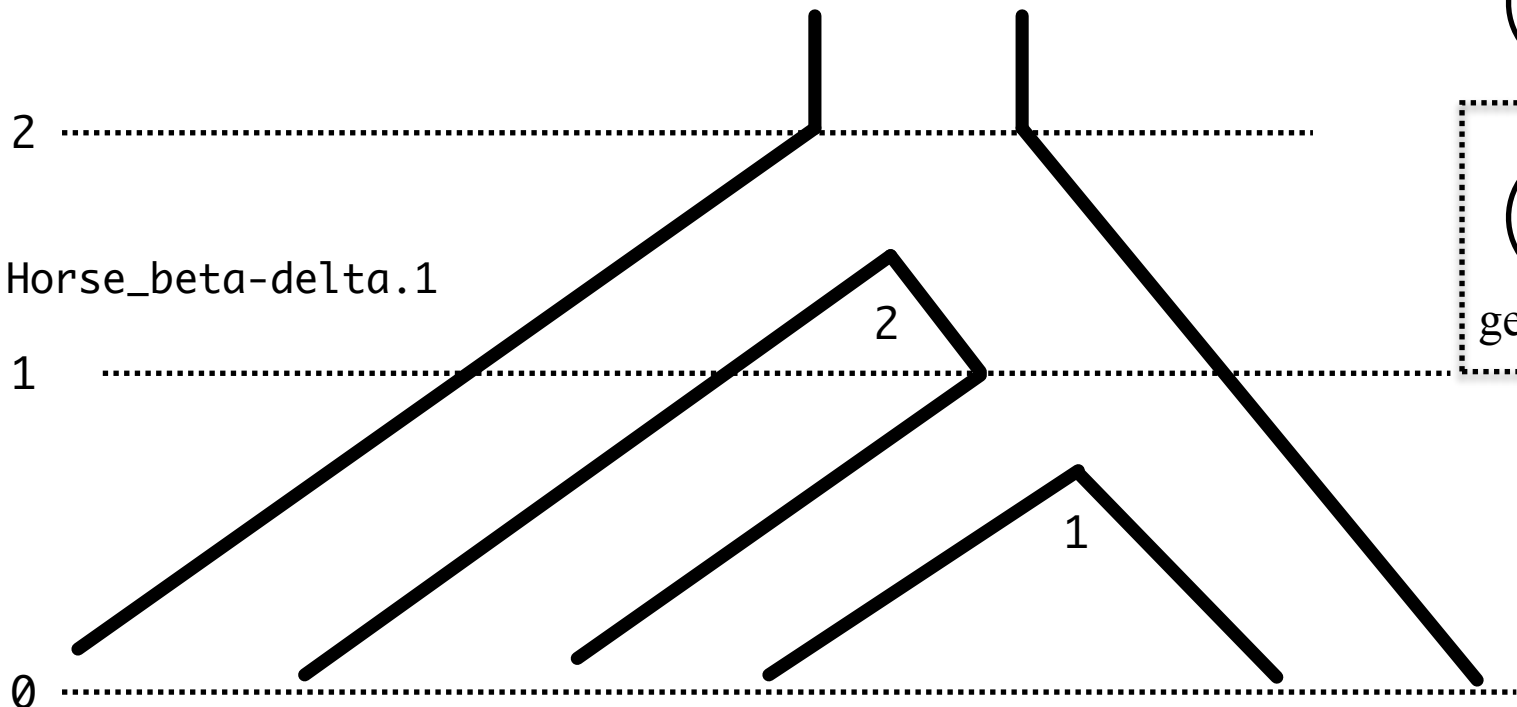
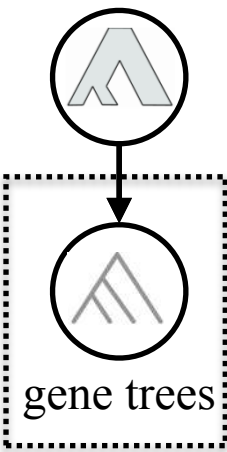
((Human_gamma:5,(Cow_beta-delta:1,Cow_gamma:1).1D@0|0.15|Cow:2).2:1,(Horse_beta-delta.1:2,(Human_beta:1,Human_delta:1)D@0|0.05|Human:1).2:1)D@2|1.3|2:0;

# of	Duplications	Transfers	Losses	Speciations
Total	3	0	2	4

# of	Duplications	Transfers	Losses	copies	
S_terminal_branch	Cow	1	0	1	2
S_terminal_branch	Horse	0	0	1	1
S_terminal_branch	Human	1	0	0	3
S_internal_branch	1	0	0	0	2
S_internal_branch	2	1	0	0	2

Zukerkandl-Pauling.tree.ale.ml_rec (END)

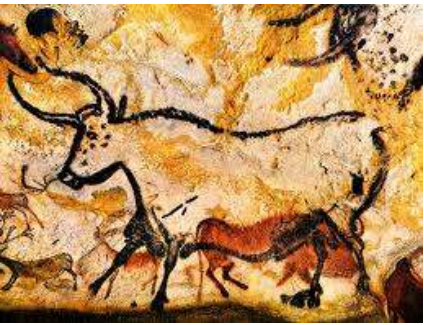
DL
species tree



Human



Horse



Cow

view the tree using either
phylo.io, seaview or FigTree!

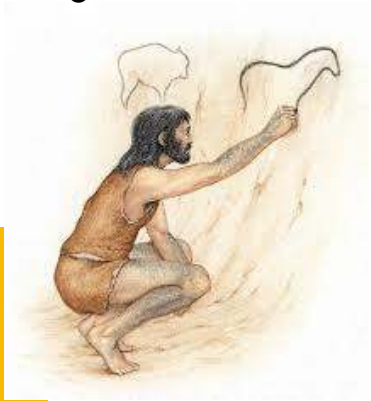
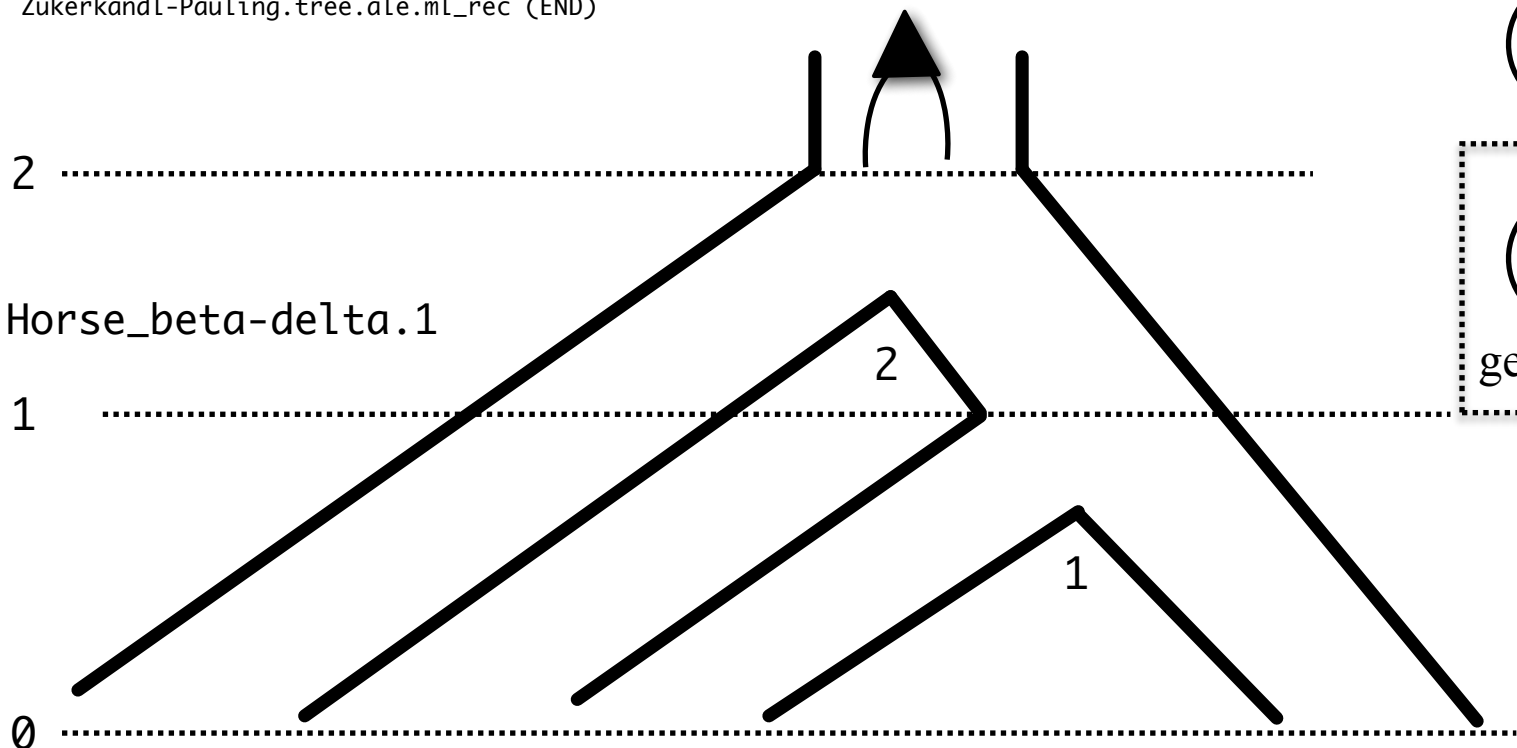
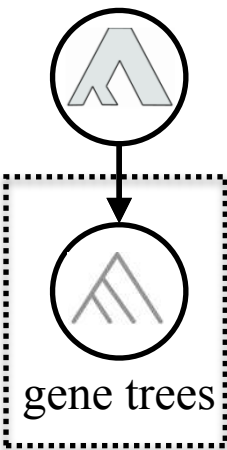
Hemoglobin/

ALEmI HuHoCo.tree Zukerkandl-Pauling.tree.ale **delta=0.01 lambda=0.01** tau=0 sample=10

((Human_gamma:5,(Cow_beta-delta:1,Cow_gamma:1).1D@0|0.15|Cow:2).2:1,(Horse_beta-delta.1:2,(Human_beta:1,Human_delta:1)D@0|0.05|Human:1).2:1)D@2|1.3|2:0;

# of Total	Duplications 3	Transfers 0	Losses 2	Speciations 4
# of S_terminal_branch	Duplications Cow	Transfers 1	Losses 0	copies 1 2
S_terminal_branch	Horse	0	0	1 1
S_terminal_branch	Human	1	0	0 3
S_internal_branch	1	0	0	0 2
S_internal_branch	2	1	0	0 2
Zukerkandl-Pauling.tree.ale.ml_rec (END)				

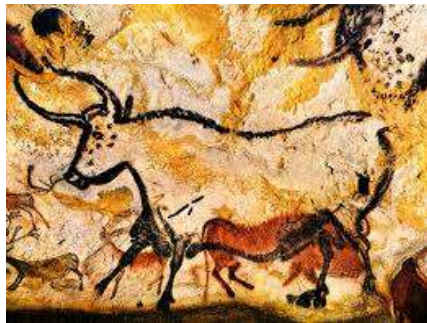
DL
species tree



Human



Horse



Cow

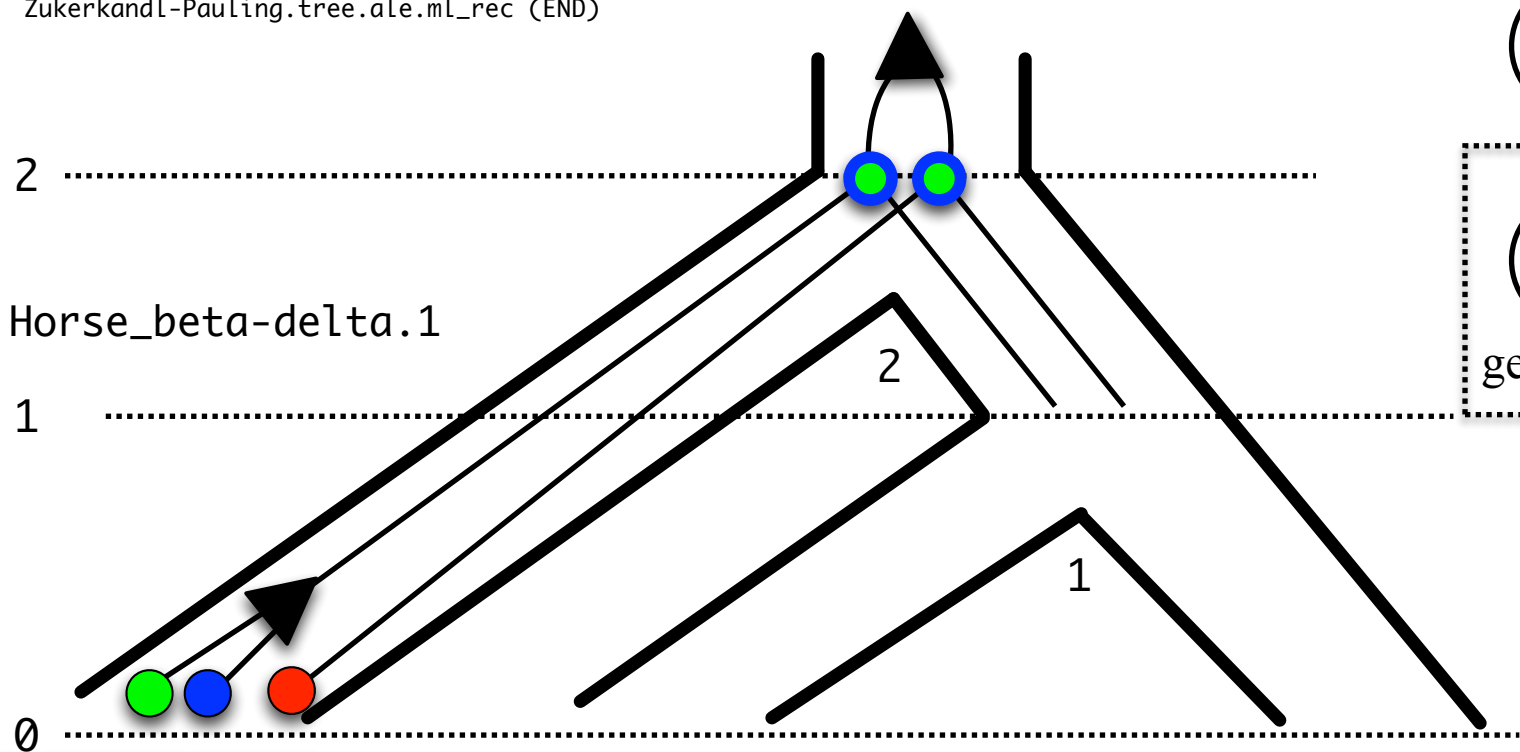
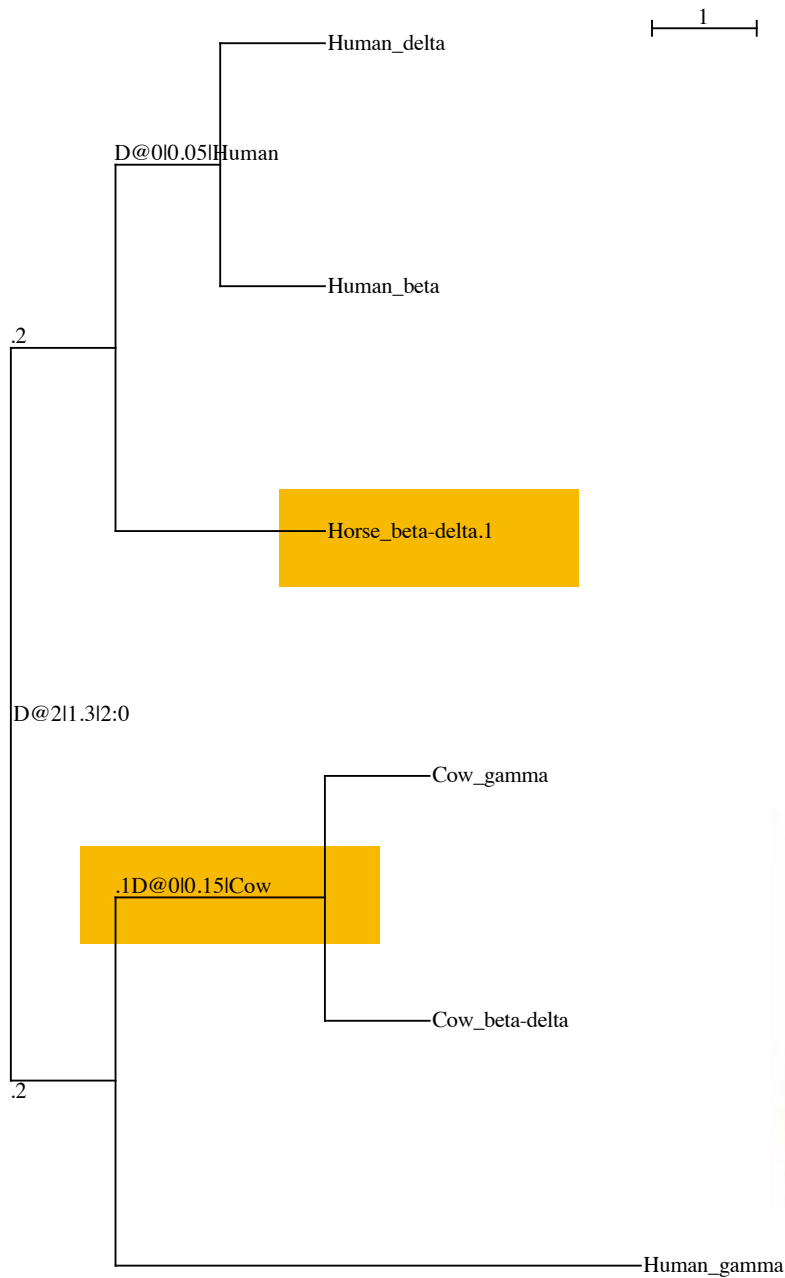
view the tree using either
phylo.io, seaview or FigTree!

Hemoglobin/

ALEmI HuHoCo.tree Zukerkandl-Pauling.tree.ale **delta=0.01 lambda=0.01** tau=0 sample=10

((Human_gamma:5,(Cow_beta-delta:1,Cow_gamma:1).1D@0|0.15|Cow:2).2:1,(Horse_beta-delta.1:2,(Human_beta:1,Human_delta:1)D@0|0.05|Human:1).2:1)D@2|1.3|2:0;

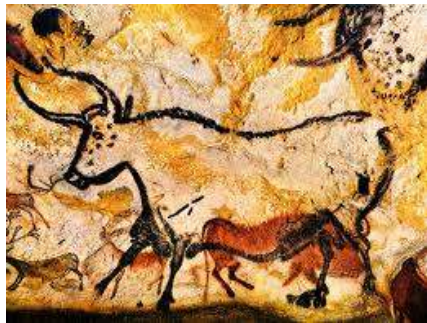
# of Total	Duplications 3	Transfers 0	Losses 2	Speciations 4
# of S_terminal_branch	Duplications Cow	Transfers 1	Losses 0	copies 1 2
S_terminal_branch	Horse	0	0	1 1
S_terminal_branch	Human	1	0	0 3
S_internal_branch	1	0	0	0 2
S_internal_branch	2	1	0	0 2
Zukerkandl-Pauling.tree.ale.ml_rec (END)				



Human

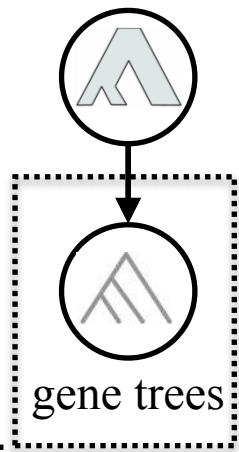


Horse



Cow

DL
species tree



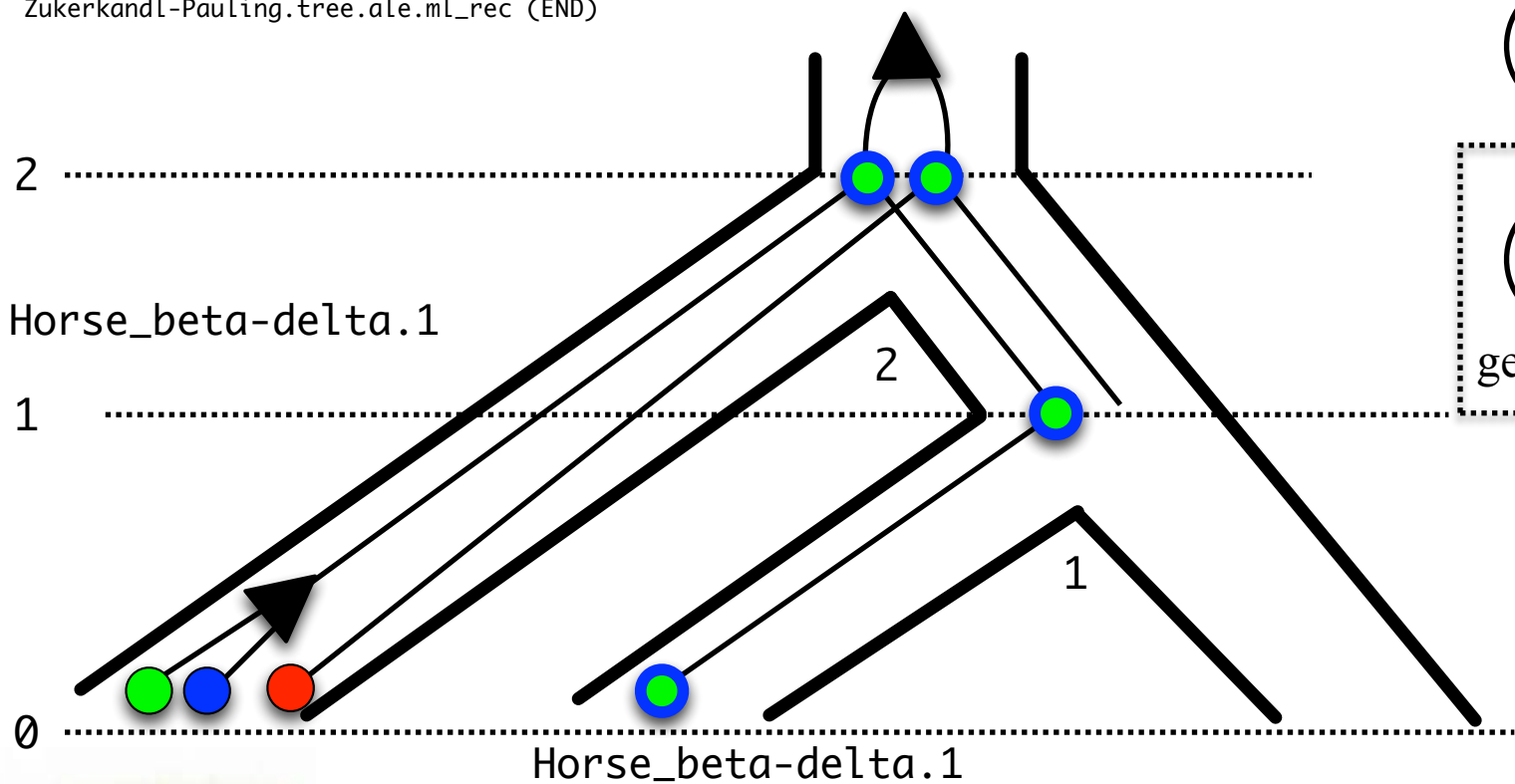
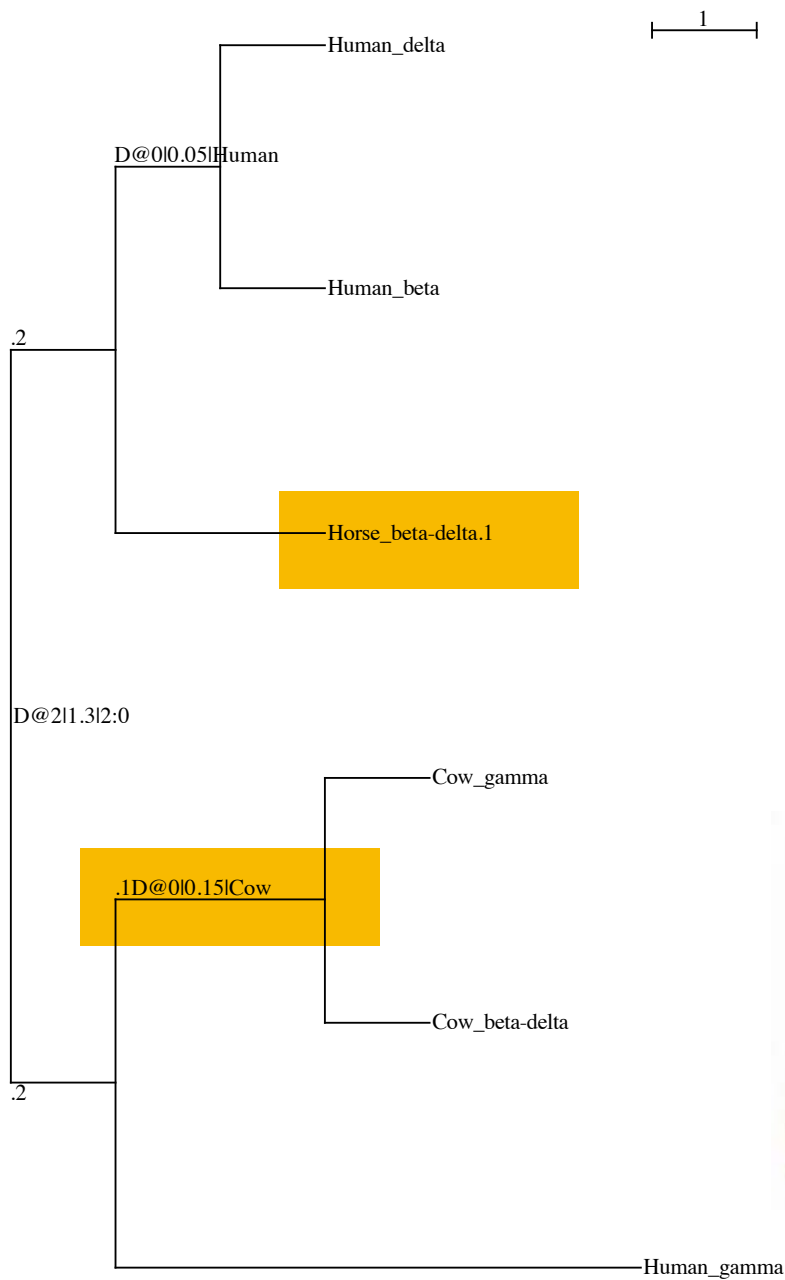
view the tree using either
phylo.io, seaview or FigTree!

Hemoglobin/

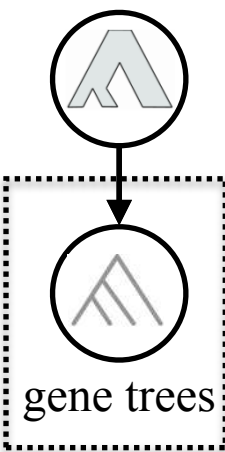
ALEmI HuHoCo.tree Zukerkandl-Pauling.tree.ale **delta=0.01 lambda=0.01** tau=0 sample=10

((Human_gamma:5,(Cow_beta-delta:1,Cow_gamma:1).1D@0|0.15|Cow:2).2:1,(Horse_beta-delta.1:2,(Human_beta:1,Human_delta:1)D@0|0.05|Human:1).2:1)D@2|1.3|2:0;

# of Total	Duplications 3	Transfers 0	Losses 2	Speciations 4	
# of	Duplications	Transfers	Losses	copies	
S_terminal_branch	Cow	1	0	1	2
S_terminal_branch	Horse	0	0	1	1
S_terminal_branch	Human	1	0	0	3
S_internal_branch	1	0	0	0	2
S_internal_branch	2	1	0	0	2
Zukerkandl-Pauling.tree.ale.ml_rec (END)					



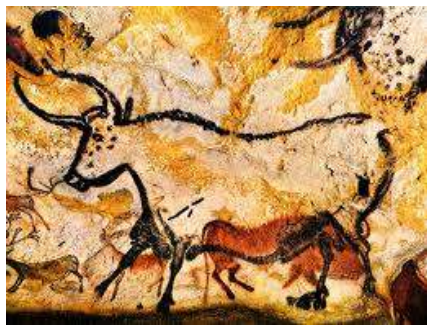
DL species tree



Human



Horse



Cow

view the tree using either
phylo.io, seaview or FigTree!

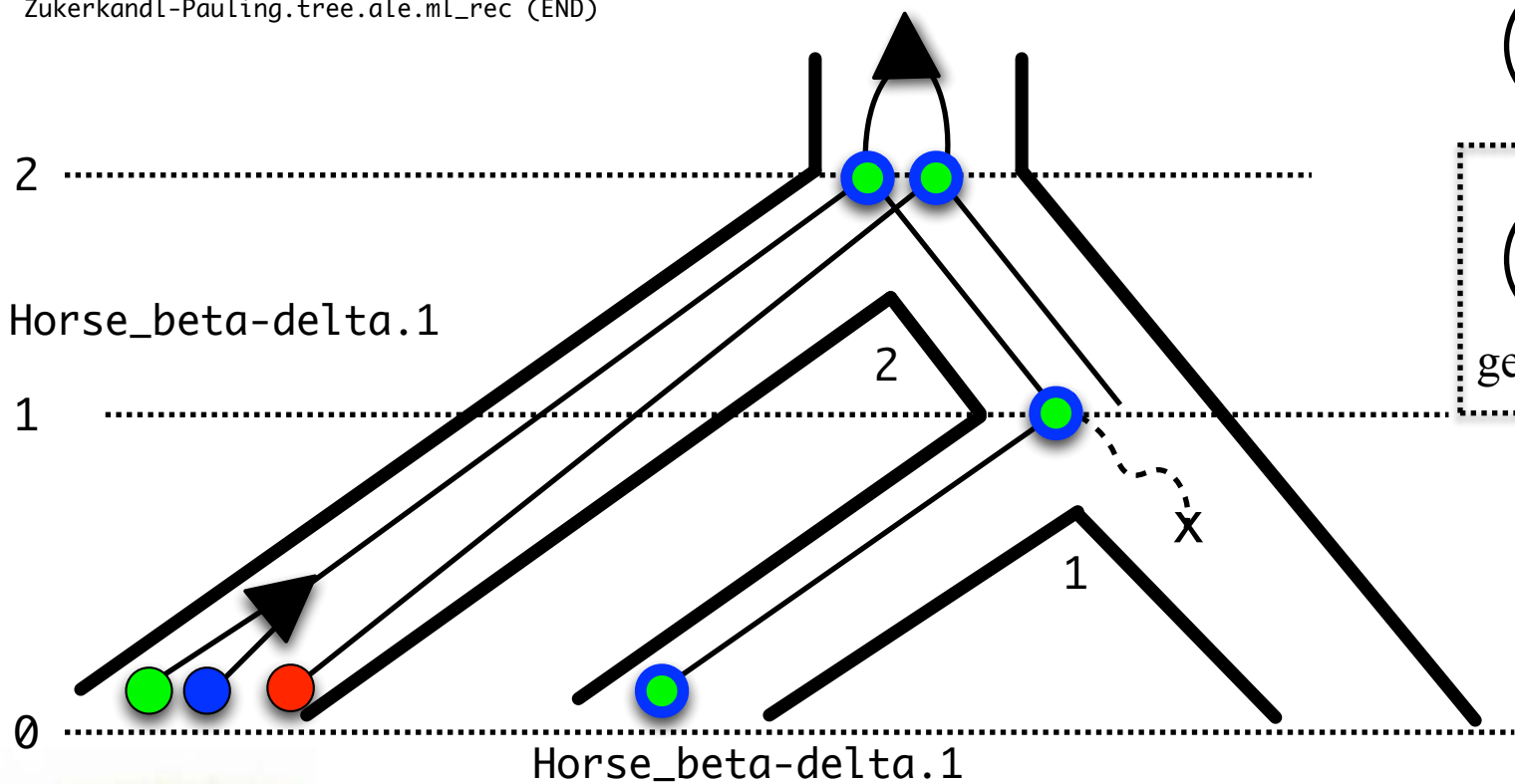
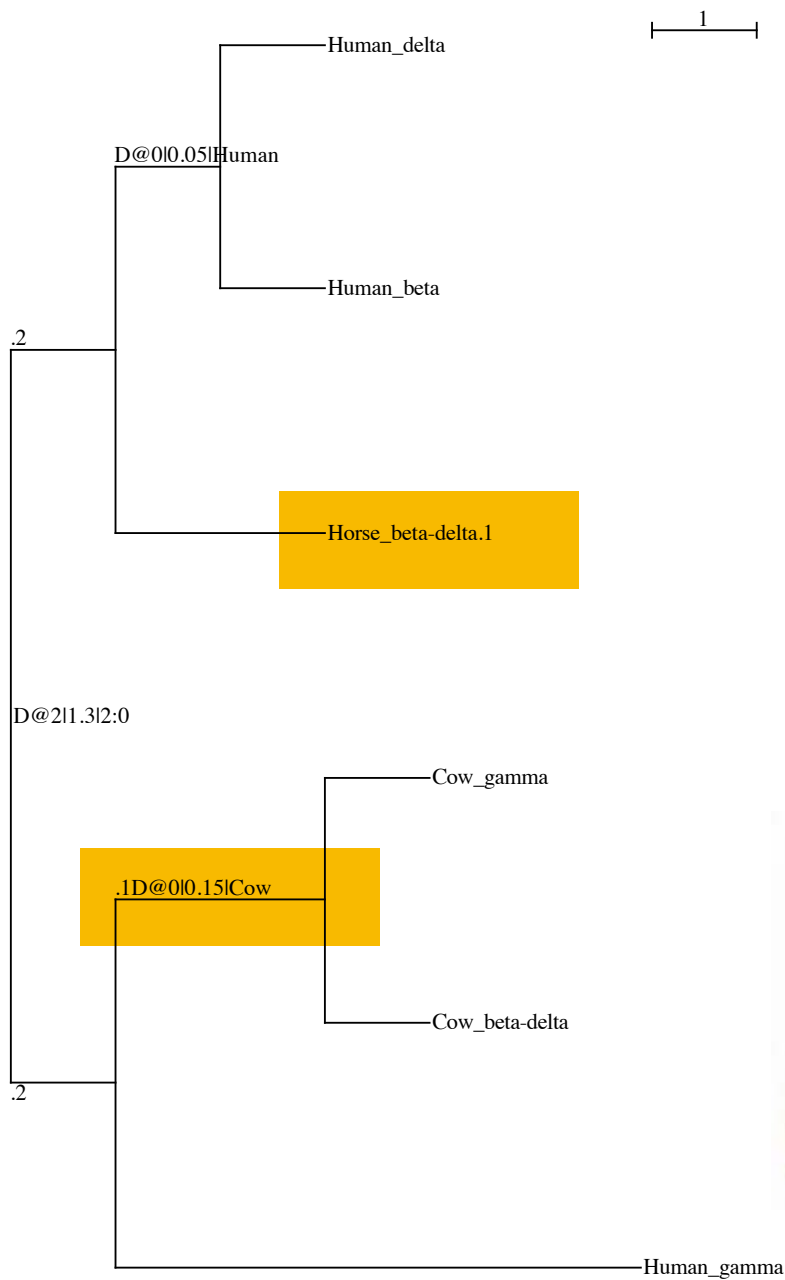
Hemoglobin/

ALEmI HuHoCo.tree Zukerkandl-Pauling.tree.ale **delta=0.01 lambda=0.01** tau=0 sample=10

((Human_gamma:5,(Cow_beta-delta:1,Cow_gamma:1).1D@0|0.15|Cow:2).2:1,(Horse_beta-delta.1:2,(Human_beta:1,Human_delta:1)D@0|0.05|Human:1).2:1)D@2|1.3|2:0;

# of Total	Duplications 3	Transfers 0	Losses 2	Speciations 4
# of S_terminal_branch	Duplications Cow 1	Transfers 1	Losses 0	copies 1 2
S_terminal_branch	Horse 0	0	0	1 1
S_terminal_branch	Human 1	1	0	0 3
S_internal_branch	1 0	0	0	0 2
S_internal_branch	2 1	0	0	0 2

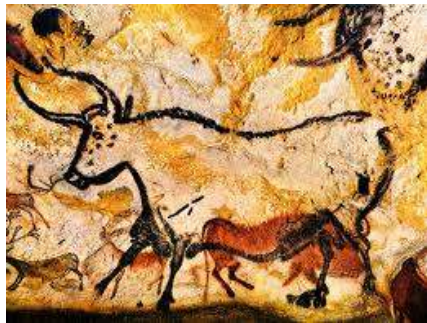
Zukerkandl-Pauling.tree.ale.ml_rec (END)



Human



Horse



Cow

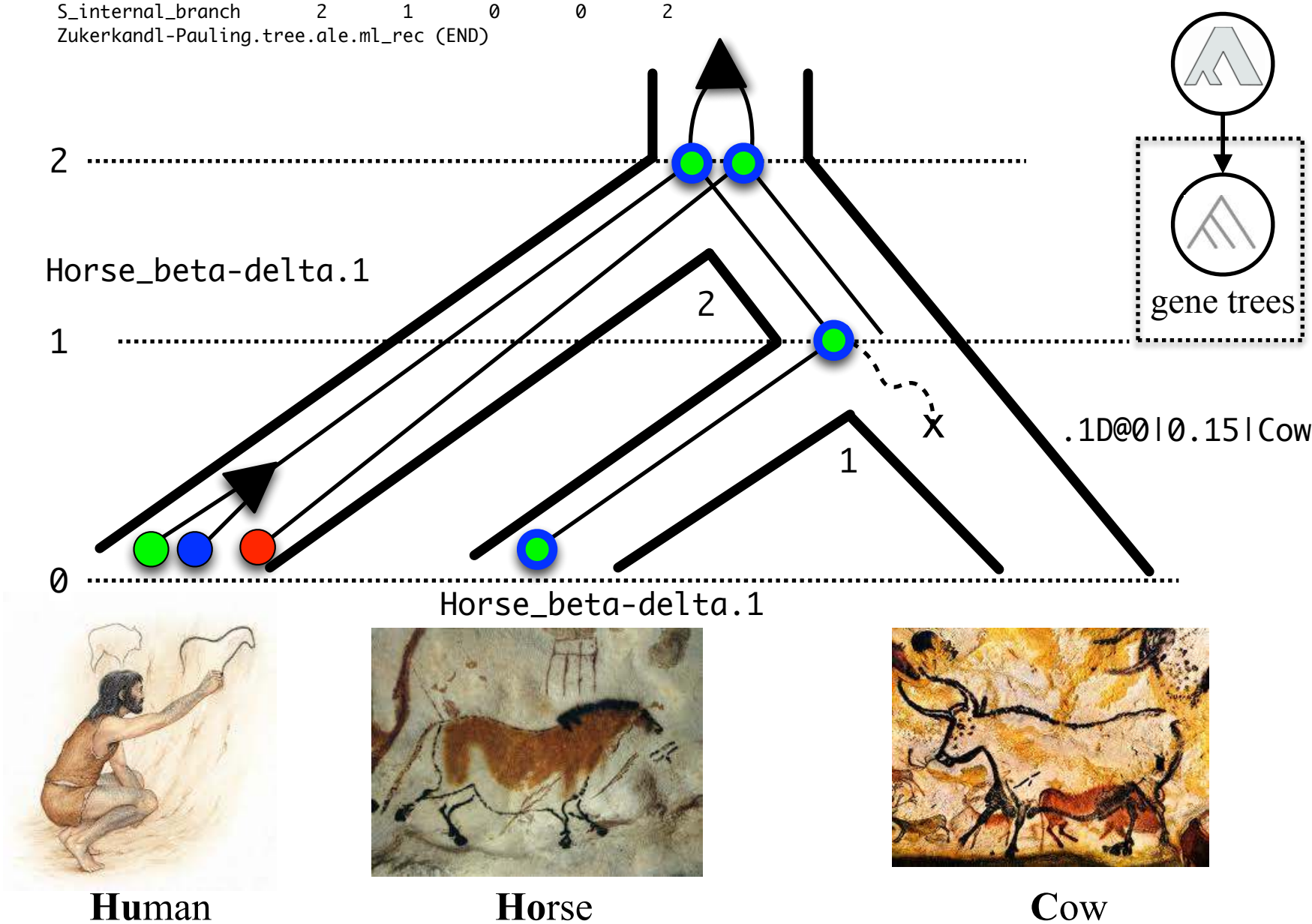
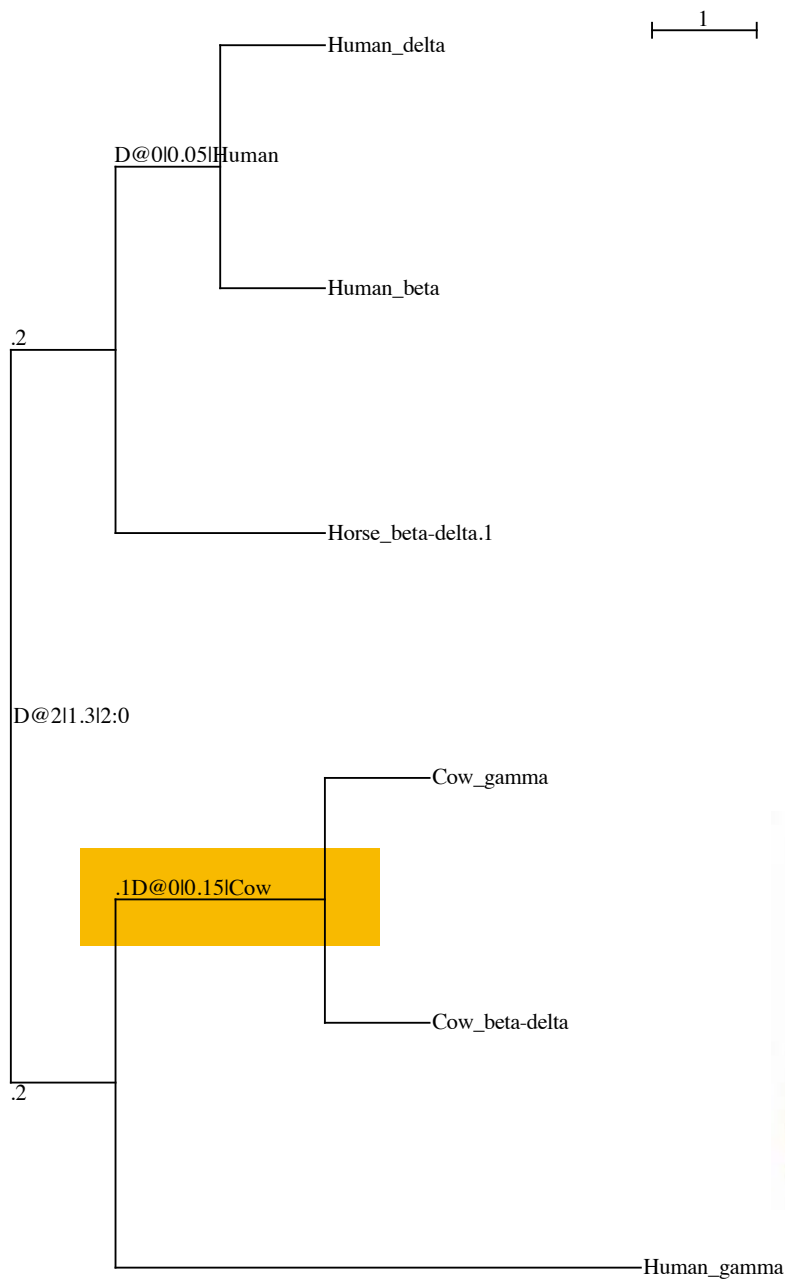
view the tree using either
phylo.io, seaview or FigTree!

Hemoglobin/

ALEmI HuHoCo.tree Zukerkandl-Pauling.tree.ale **delta=0.01 lambda=0.01** tau=0 sample=10

((Human_gamma:5,(Cow_beta-delta:1,Cow_gamma:1).1D@0|0.15|Cow:2).2:1,(Horse_beta-delta.1:2,(Human_beta:1,Human_delta:1)D@0|0.05|Human:1).2:1)D@2|1.3|2:0;

# of Total	Duplications 3	Transfers 0	Losses 2	Speciations 4	
# of	Duplications	Transfers	Losses	copies	
S_terminal_branch	Cow	1	0	1	2
S_terminal_branch	Horse	0	0	1	1
S_terminal_branch	Human	1	0	0	3
S_internal_branch	1	0	0	0	2
S_internal_branch	2	1	0	0	2
Zukerkandl-Pauling.tree.ale.ml_rec (END)					



view the tree using either
phylo.io, seaview or FigTree!

Hemoglobin/

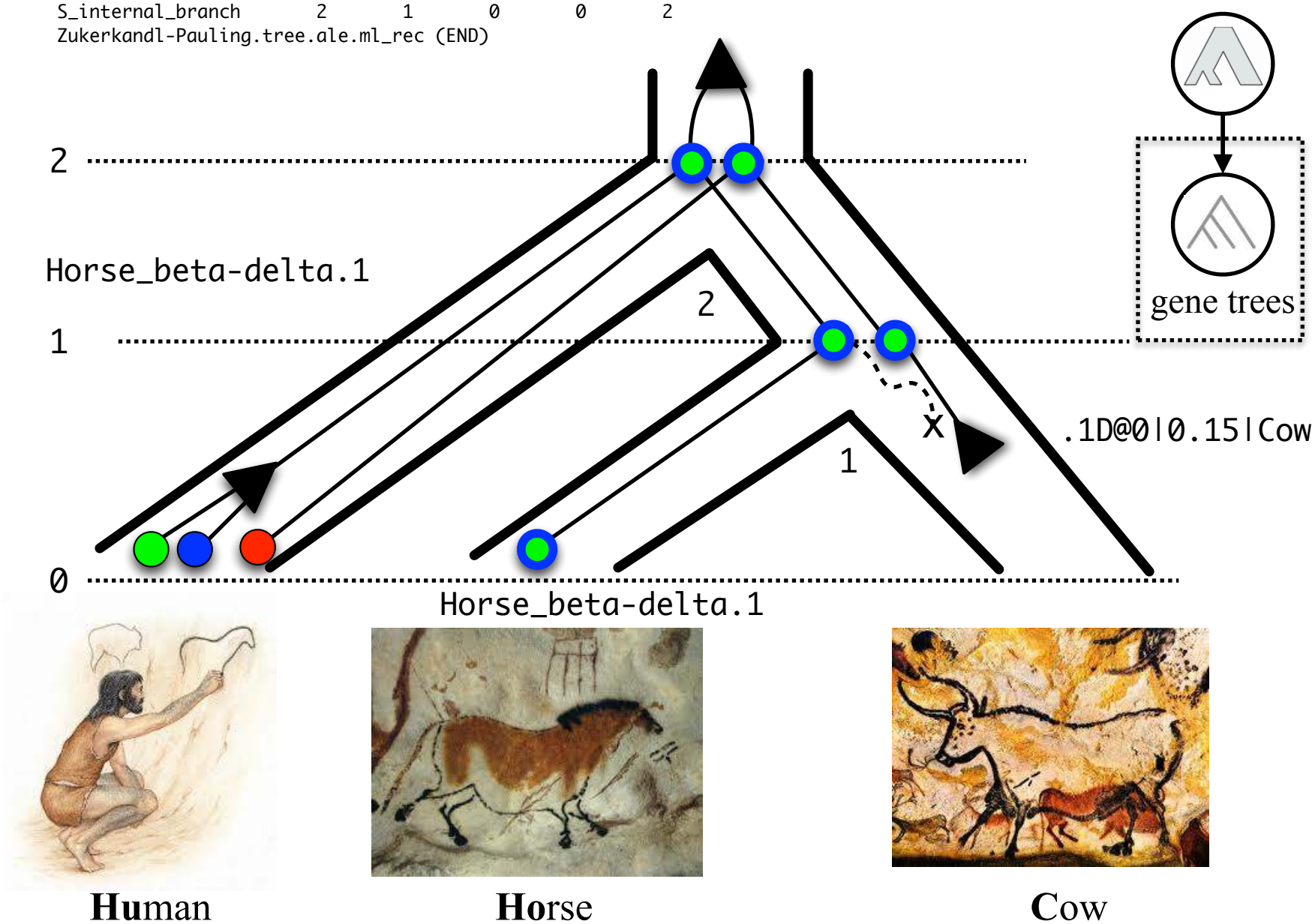
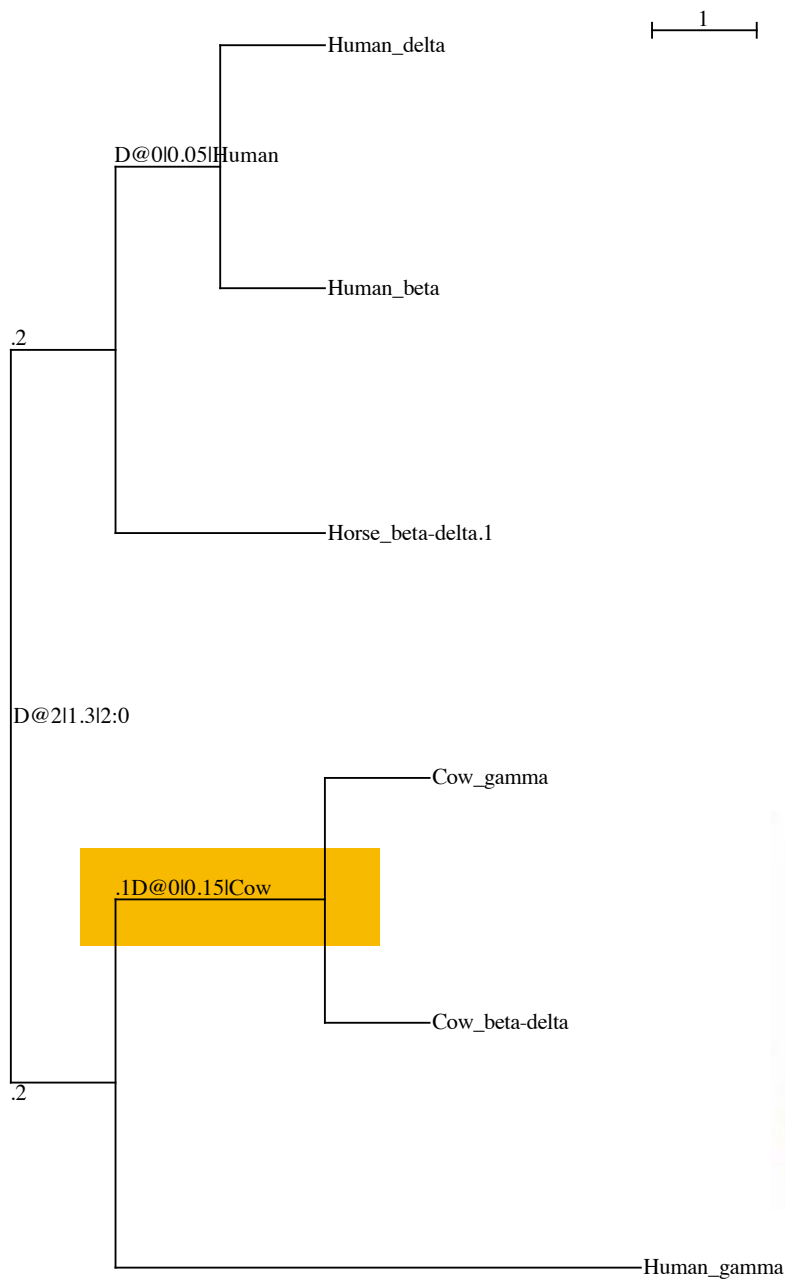
ALEmI HuHoCo.tree Zukerkandl-Pauling.tree.ale **delta=0.01 lambda=0.01** tau=0 sample=10

((Human_gamma:5,(Cow_beta-delta:1,Cow_gamma:1).1D@0|0.15|Cow:2).2:1,(Horse_beta-delta.1:2,(Human_beta:1,Human_delta:1)D@0|0.05|Human:1).2:1)D@2|1.3|2:0;

# of	Duplications	Transfers	Losses	Speciations
Total	3	0	2	4

# of	Duplications	Transfers	Losses	copies	
S_terminal_branch	Cow	1	0	1	2
S_terminal_branch	Horse	0	0	1	1
S_terminal_branch	Human	1	0	0	3
S_internal_branch	1	0	0	0	2
S_internal_branch	2	1	0	0	2

Zukerkandl-Pauling.tree.ale.ml_rec (END)



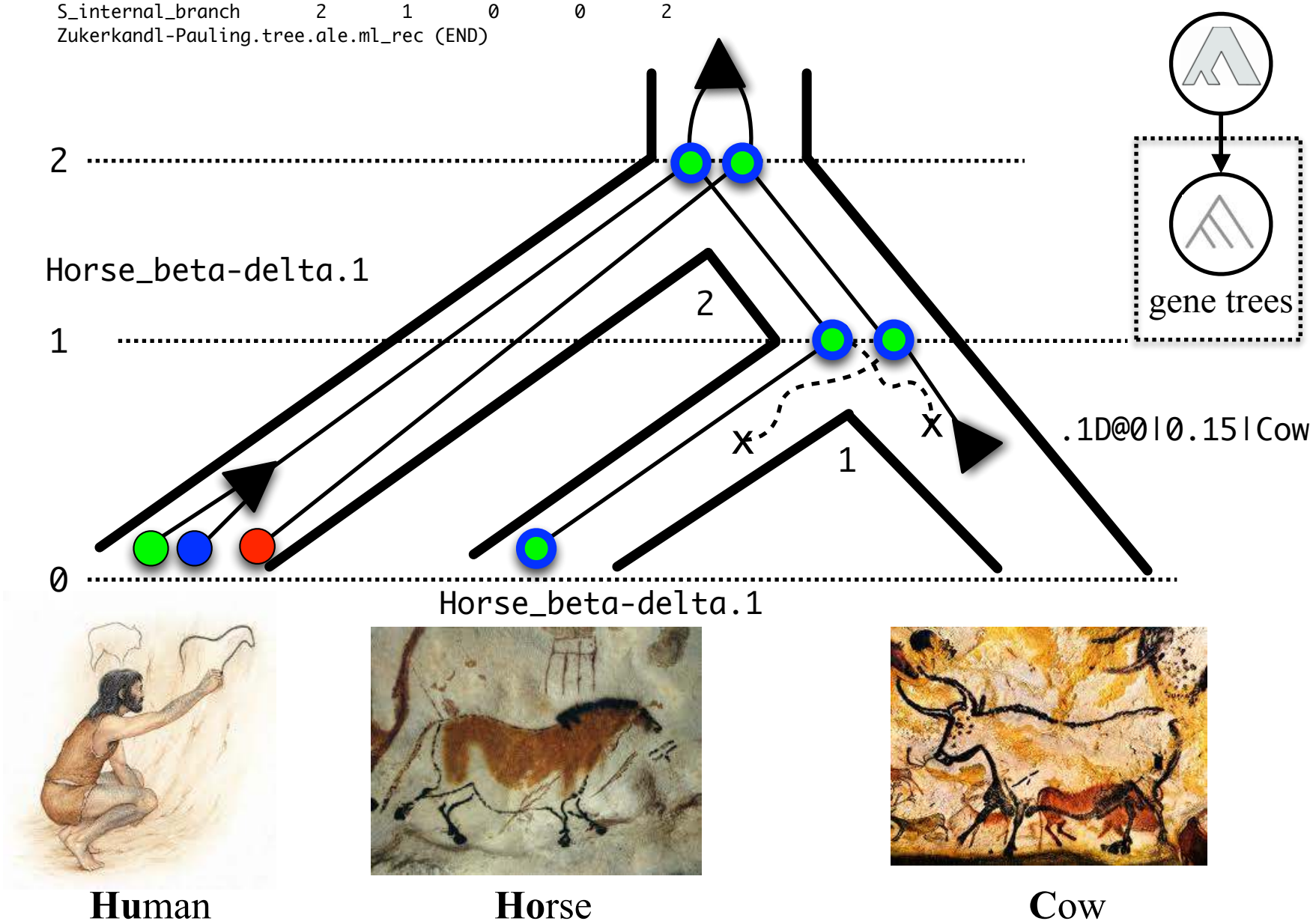
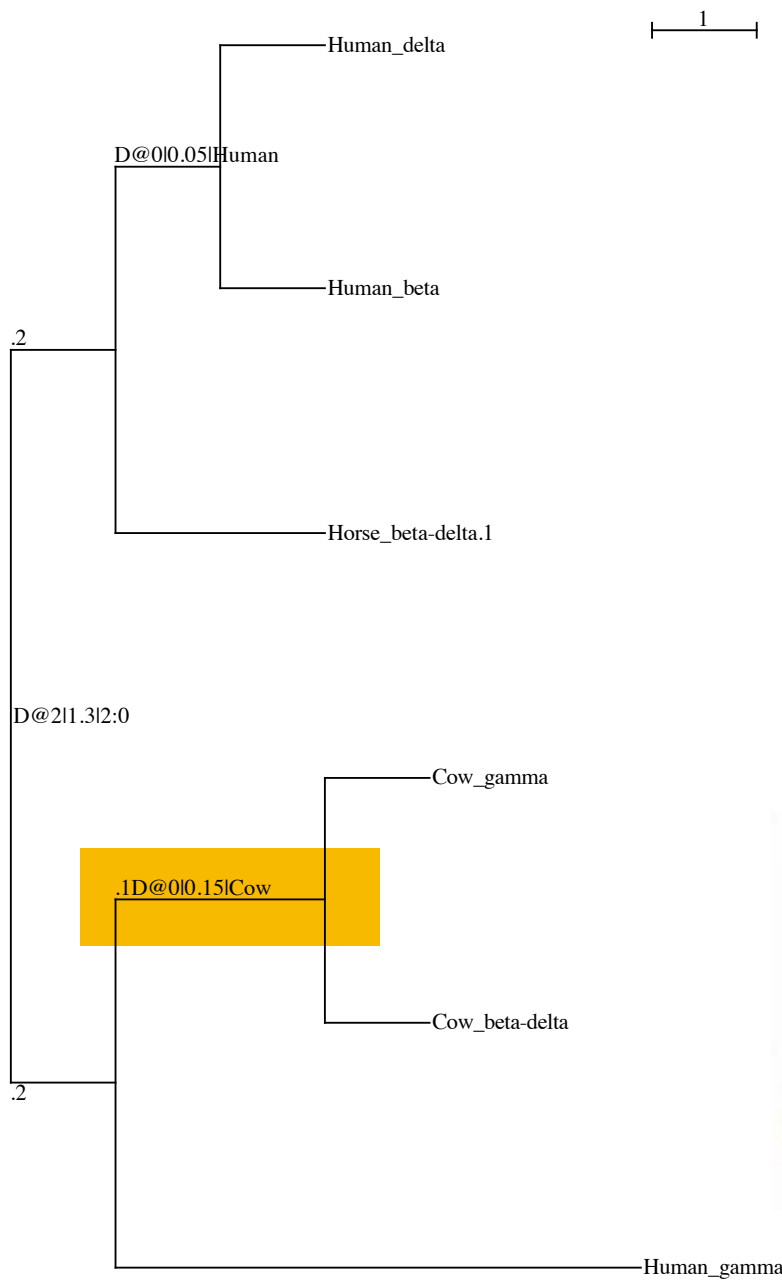
view the tree using either
phylo.io, seaview or FigTree!

Hemoglobin/

ALEmI HuHoCo.tree Zukerkandl-Pauling.tree.ale **delta=0.01 lambda=0.01** tau=0 sample=10

((Human_gamma:5,(Cow_beta-delta:1,Cow_gamma:1).1D@0|0.15|Cow:2).2:1,(Horse_beta-delta.1:2,(Human_beta:1,Human_delta:1)D@0|0.05|Human:1).2:1)D@2|1.3|2:0;

# of Total	Duplications 3	Transfers 0	Losses 2	Speciations 4	
# of	Duplications	Transfers	Losses	copies	
S_terminal_branch	Cow	1	0	1	2
S_terminal_branch	Horse	0	0	1	1
S_terminal_branch	Human	1	0	0	3
S_internal_branch	1	0	0	0	2
S_internal_branch	2	1	0	0	2
Zukerkandl-Pauling.tree.ale.ml_rec (END)					



view the tree using either
phylo.io, seaview or FigTree!

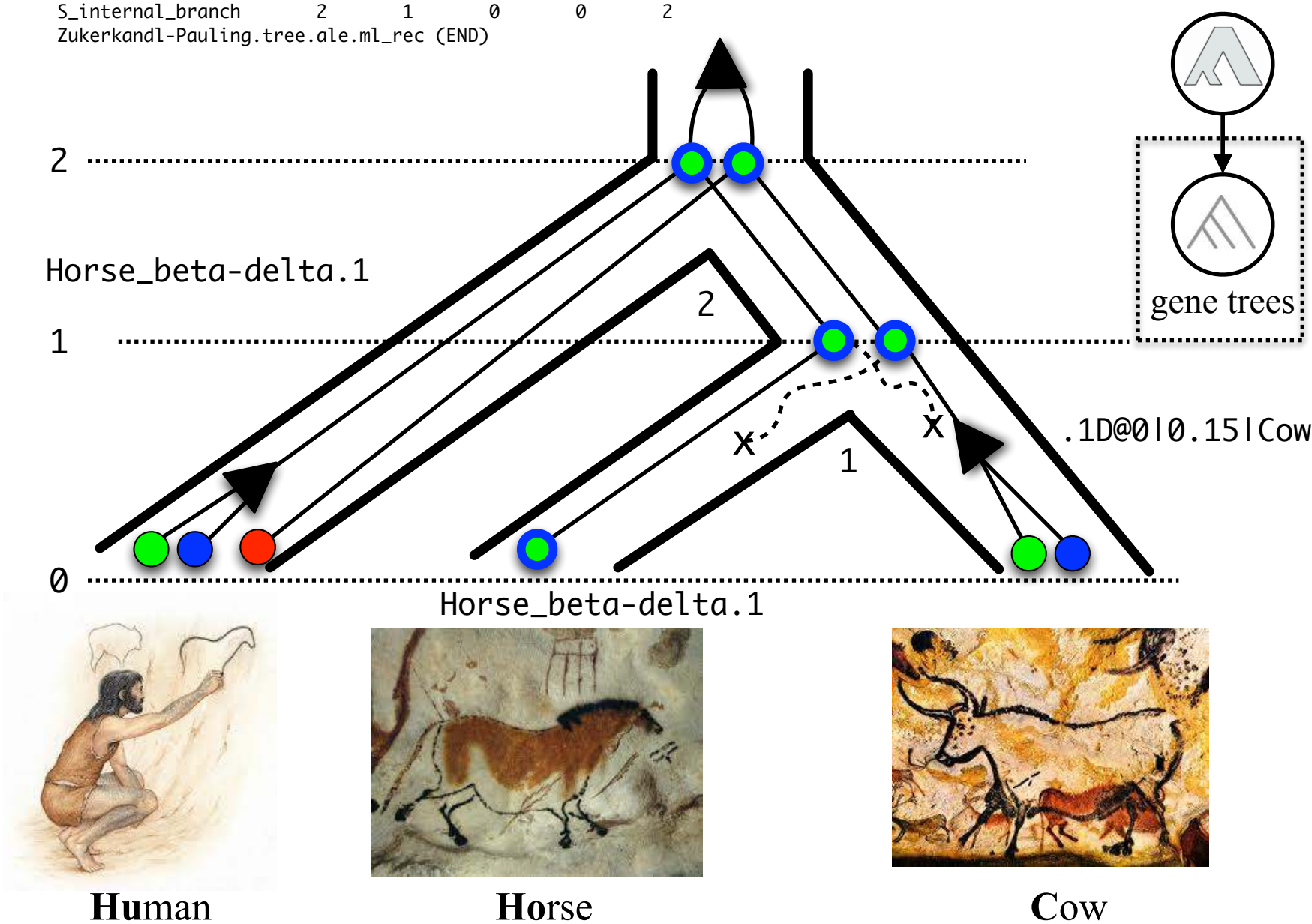
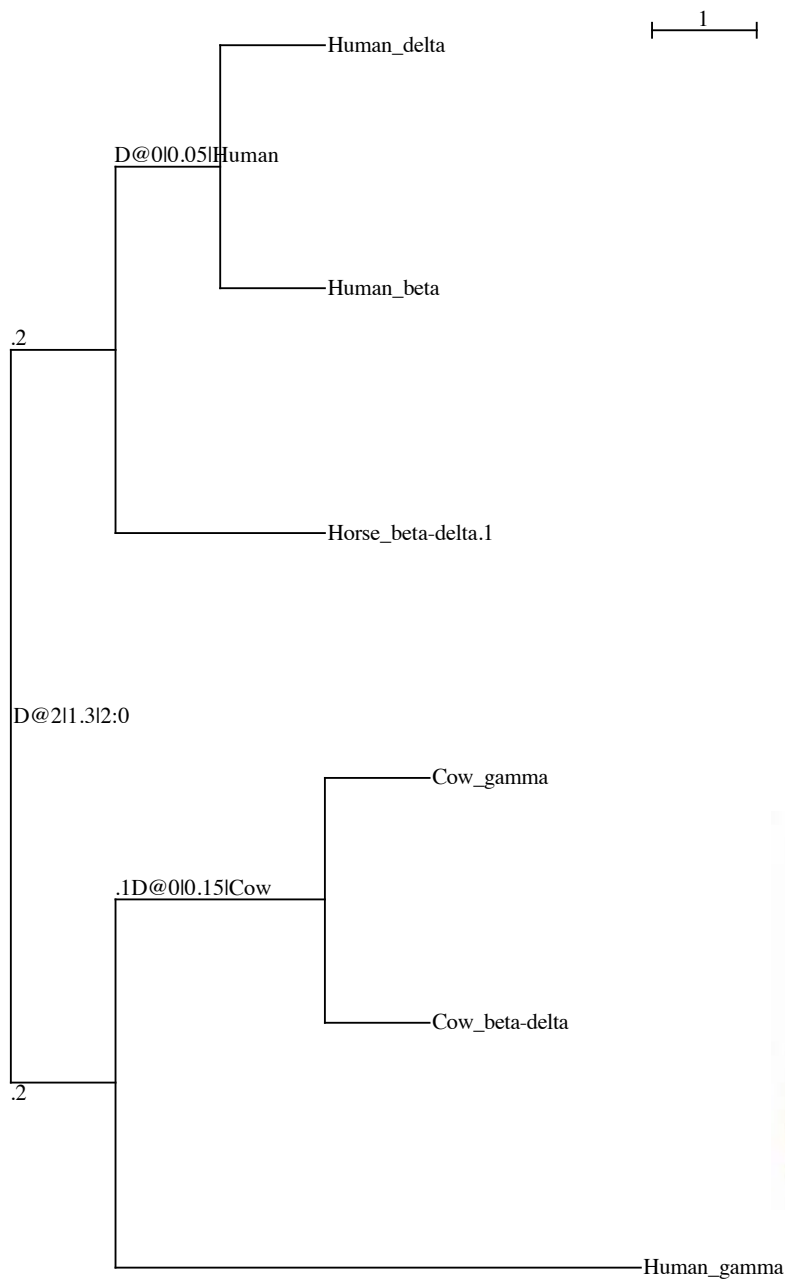
Hemoglobin/

ALEmI HuHoCo.tree Zukerkandl-Pauling.tree.ale **delta=0.01 lambda=0.01** tau=0 sample=10

((Human_gamma:5,(Cow_beta-delta:1,Cow_gamma:1).1D@0|0.15|Cow:2).2:1,(Horse_beta-delta.1:2,(Human_beta:1,Human_delta:1)D@0|0.05|Human:1).2:1)D@2|1.3|2:0;

# of Total	Duplications 3	Transfers 0	Losses 2	Speciations 4
# of S_terminal_branch	Duplications Cow 1	Transfers 1	Losses 0	copies 1 2
S_terminal_branch	Horse 0	0	0	1 1
S_terminal_branch	Human 1	1	0	0 3
S_internal_branch	1 0	0	0	0 2
S_internal_branch	2 1	0	0	0 2

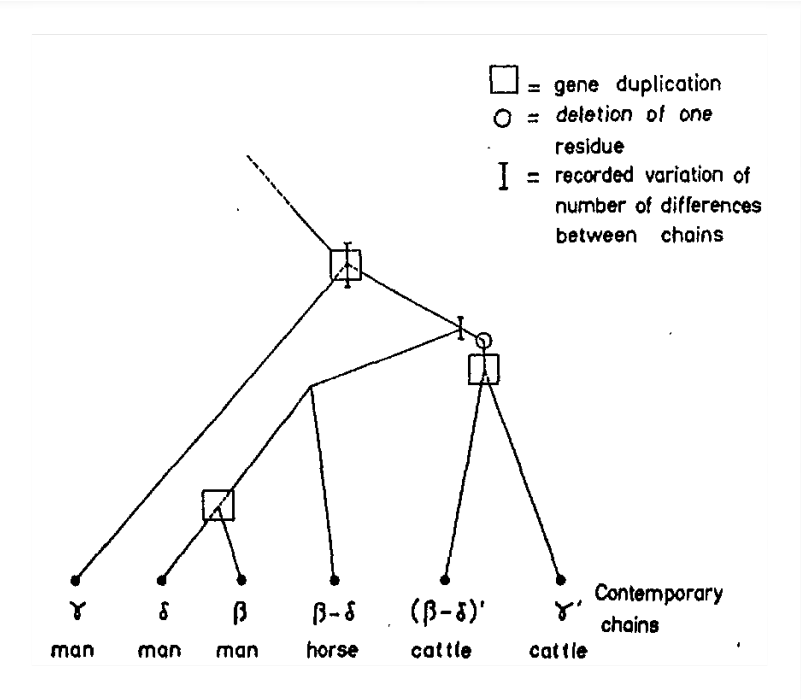
Zukerkandl-Pauling.tree.ale.ml_rec (END)



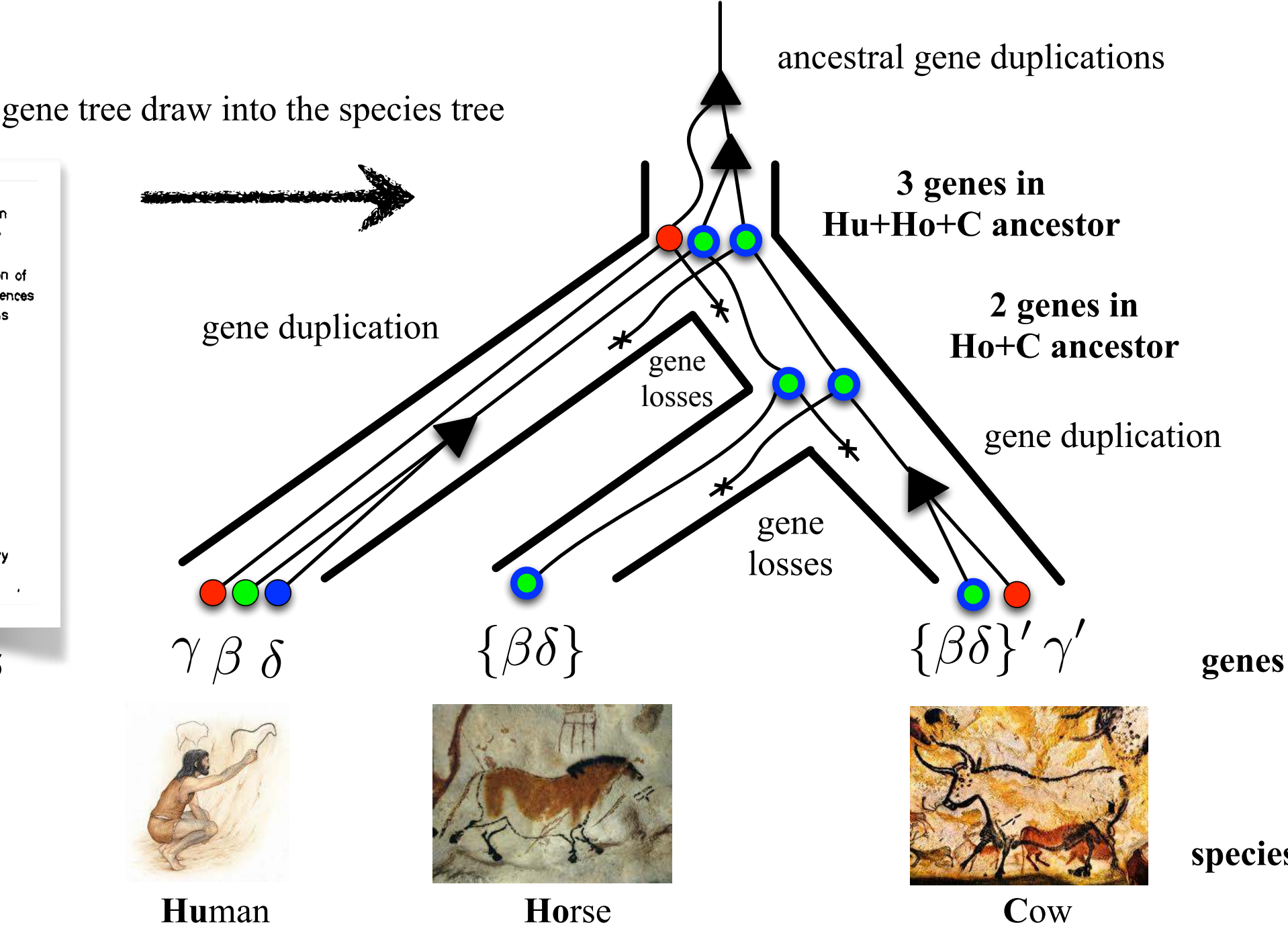
view the tree using either
phylo.io, seaview or FigTree!

The story of individual gene families is often blurred

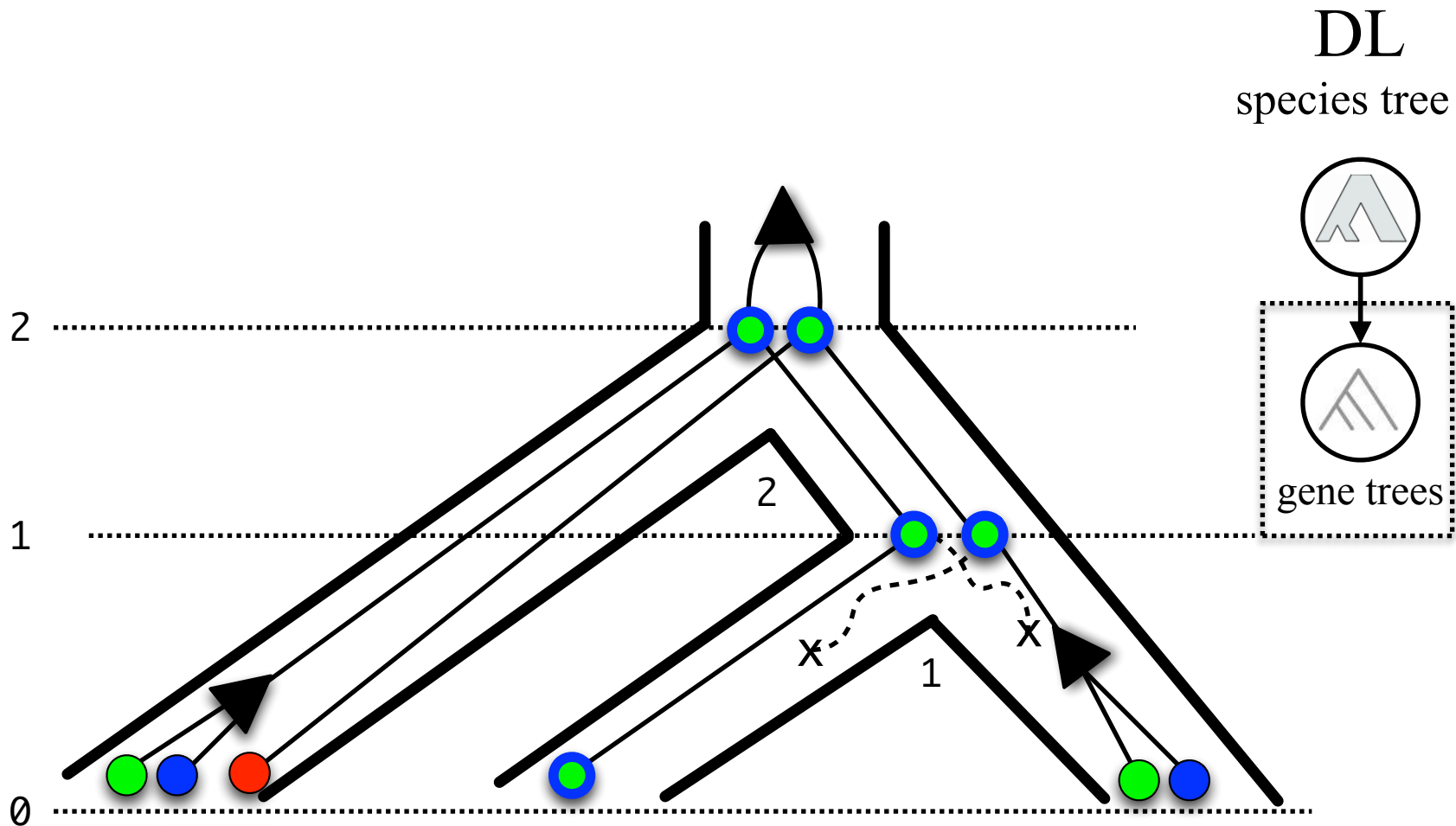
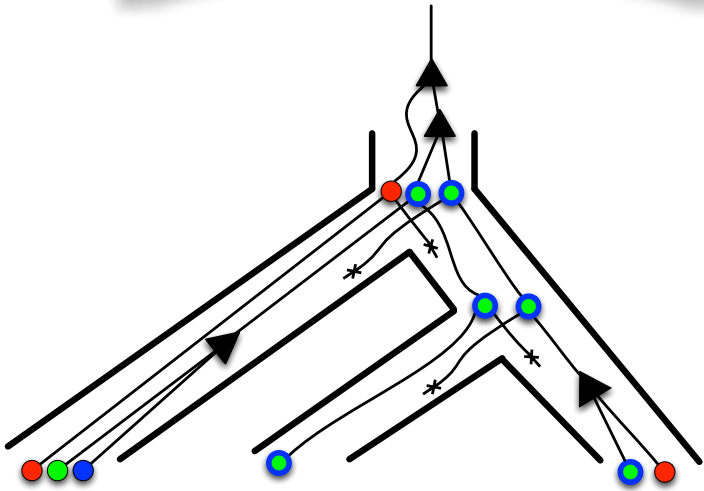
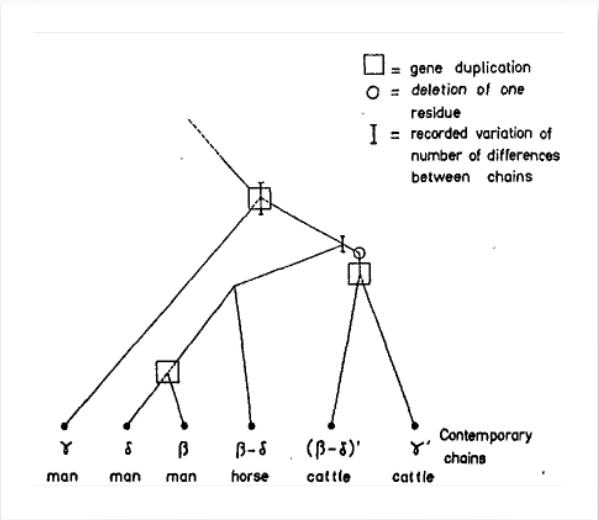
Errors in gene trees will result in conflicts with the species tree that imply spurious evolutionary events.



Zukerkandl & Pauling 1965



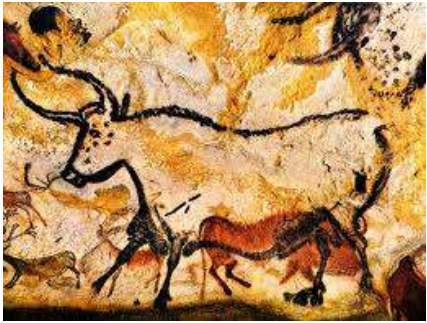
Hemoglobin/



Human



Horse

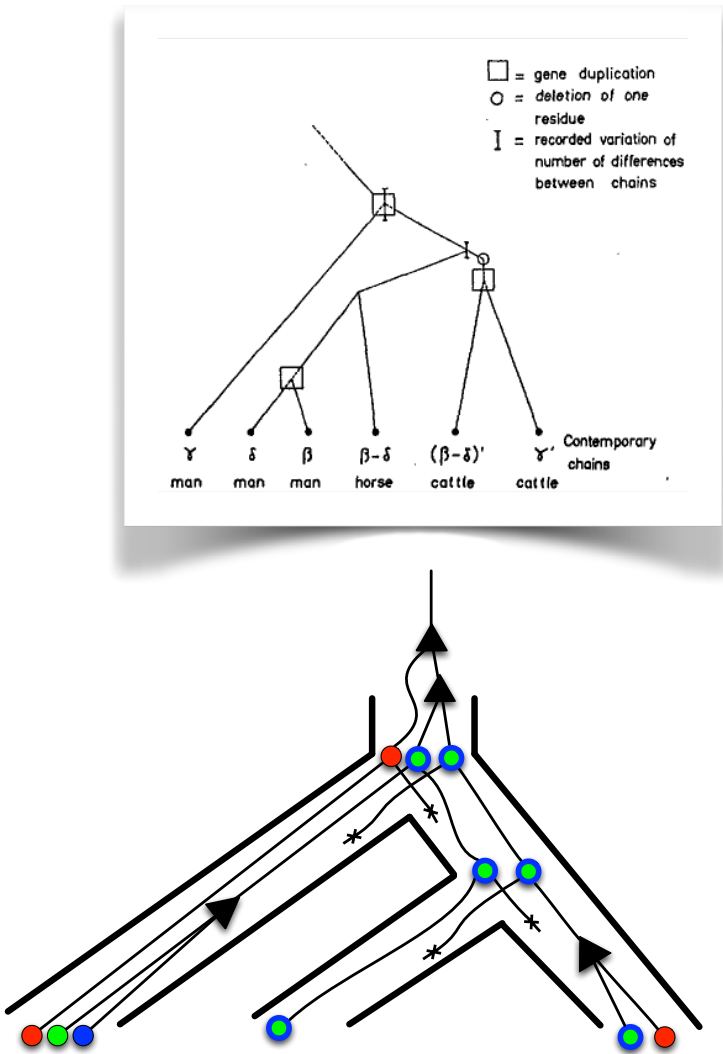


Cow

Hemoglobin/

ALeobserve Zukerkandl-Pauling_OUT.tree

ALEml **HuHoCo_OUT.tree Zukerkandl-Pauling_OUT.tree.ale** delta=0.01 lambda=0.01 tau=0 sample=10



S: ((Human:0.571429,(Horse:0.285714,Cow:0.285714)1:0.285714)2:0.428571,OUT:1)3;

Input ale from: Zukerkandl-Pauling_OUT.tree.ale

>logl: -44.1068

rate of	Duplications	Transfers	Losses
ML	0.01	0	0.01

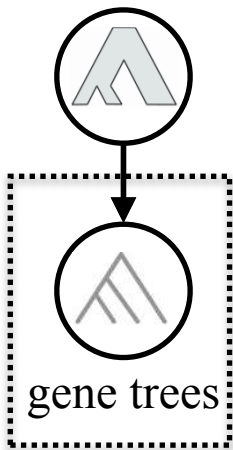
10 reconciled G-s:

# of	Duplications	Transfers	Losses	Speciations
Total	4	0	4	6

# of	Duplications	Transfers	Losses	copies	
S_terminal_branch	Cow	1	0	1	2
S_terminal_branch	Horse	0	0	1	1
S_terminal_branch	Human	1	0	1	3
S_terminal_branch	OUT	0	0	0	1
S_internal_branch	1	0	0	1	2
S_internal_branch	2	2	0	0	3
S_internal_branch	3	0	0	0	1

HuHoCo_OUT.tree_Zukerkandl-Pauling_OUT.tree.ale.ml_rec (END)

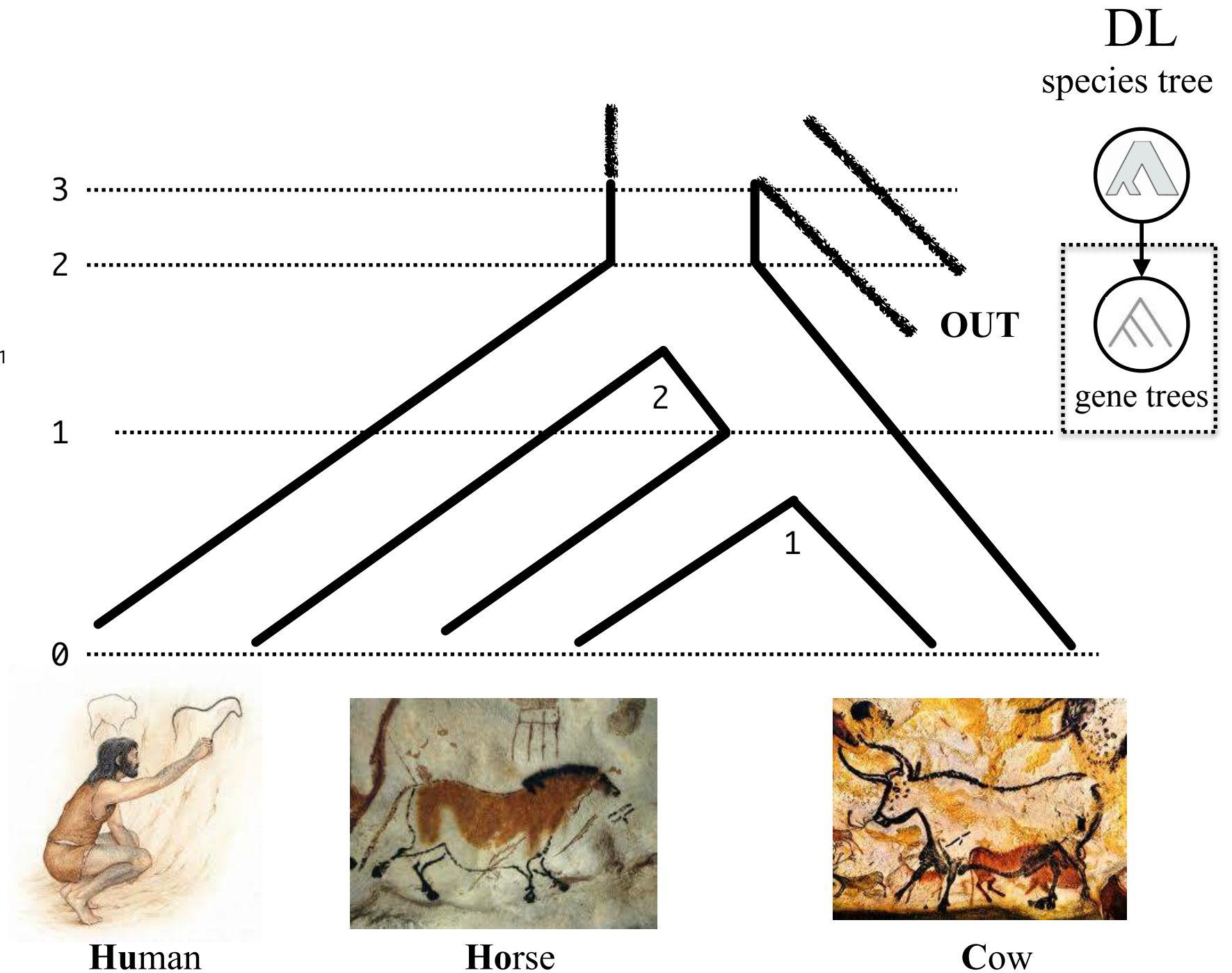
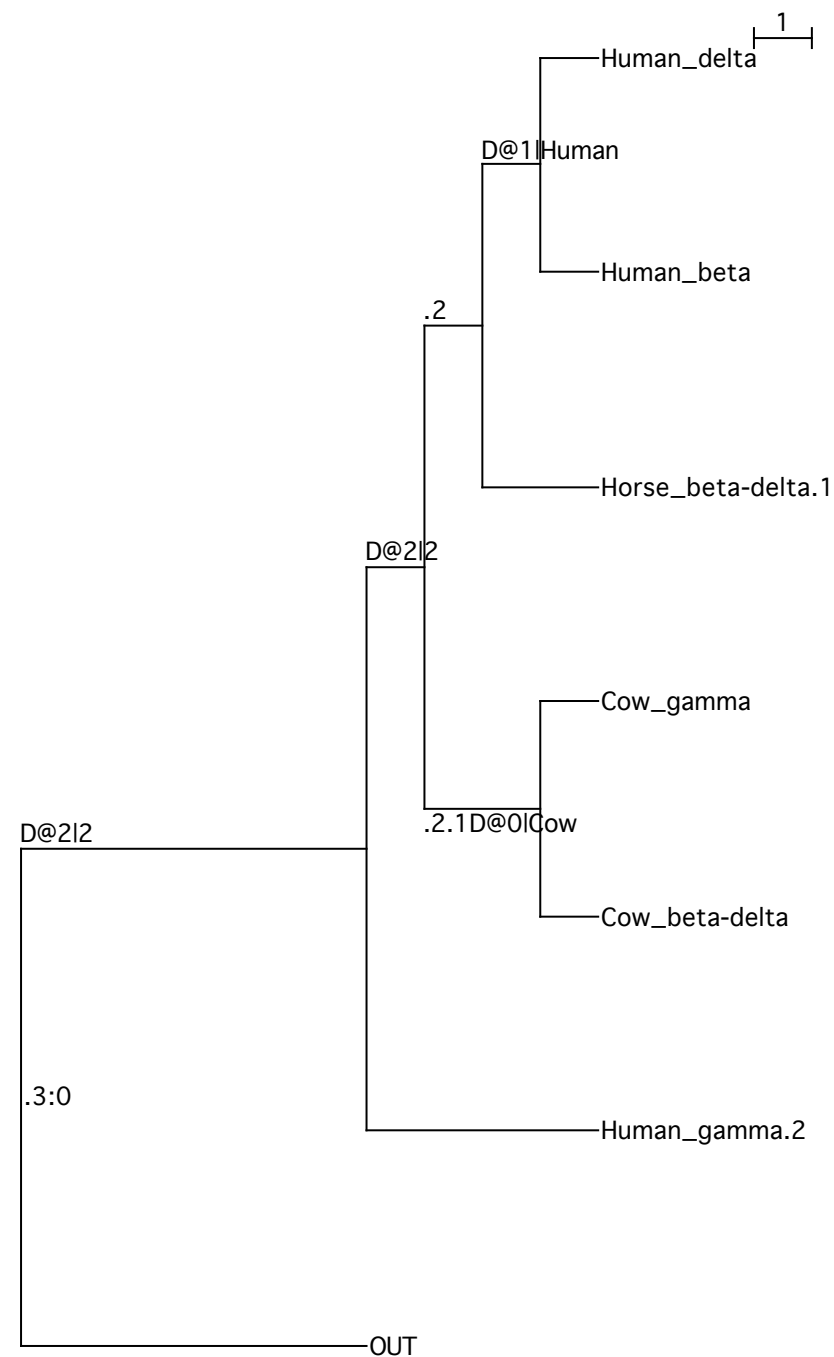
DL
species tree



Hemoglobin/

ALEmI HuHoCo_OUT.tree Zukerkandl-Pauling_OUT.tree.ale delta=0.01 lambda=0.01 tau=0 sample=10

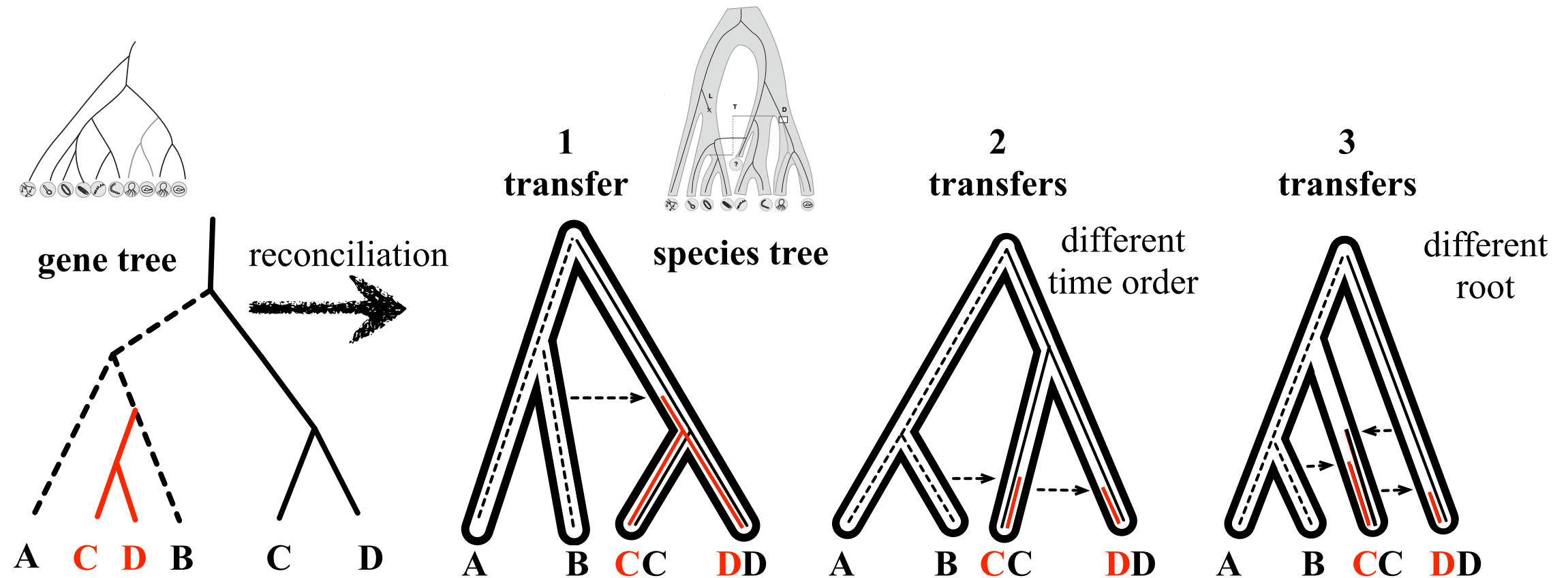
(OUT:6,(Human_gamma.2:4,((Cow_beta-delta:1,Cow_gamma:1).2.1D@0|Cow:2,(Horse_beta-delta.1:2,(Human_beta:1,Human_delta:1)D@1|Human:1).2:1)D@2|2:1)D@2|2:6).3:0;

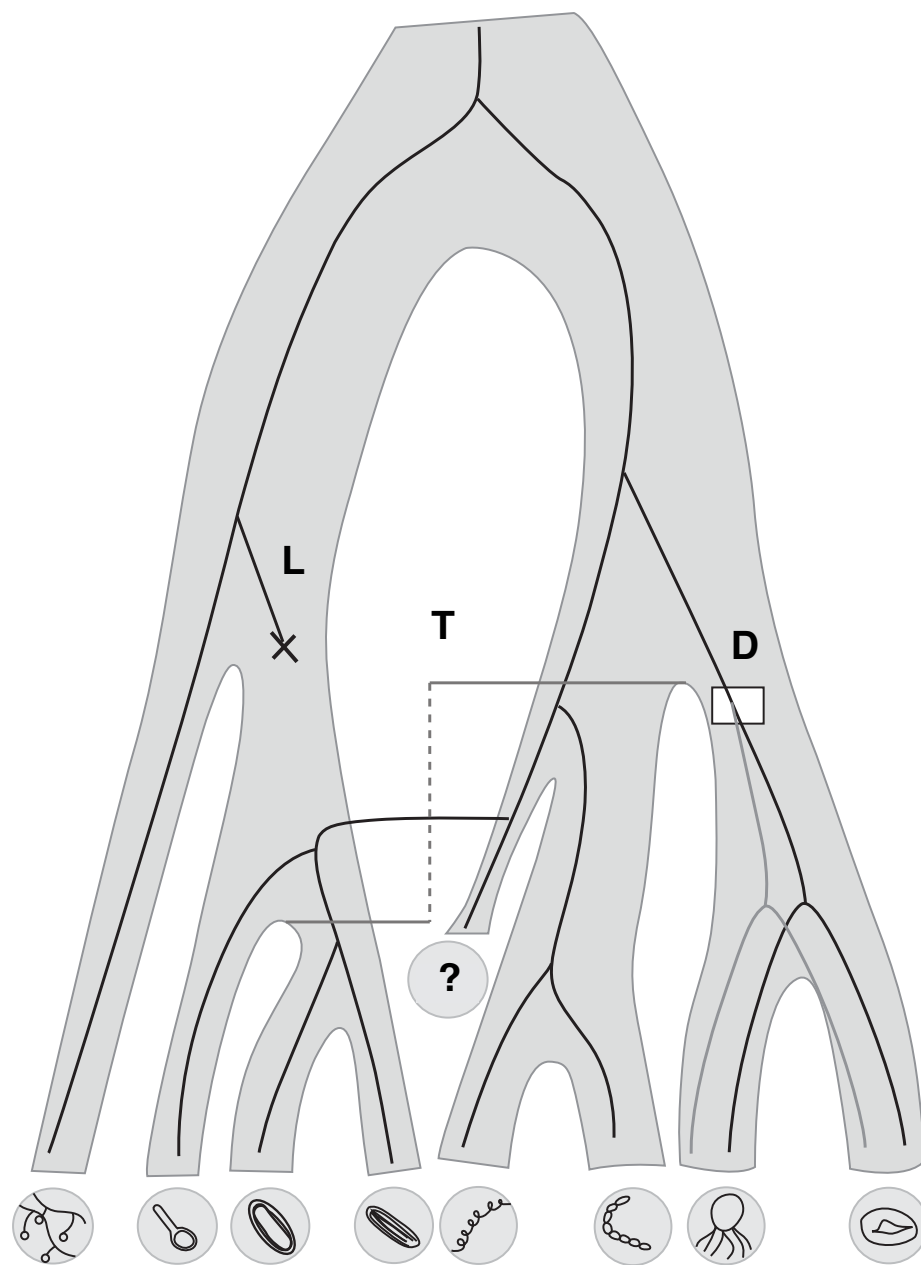


view the tree using either
phylo.io, seaview or FigTree!

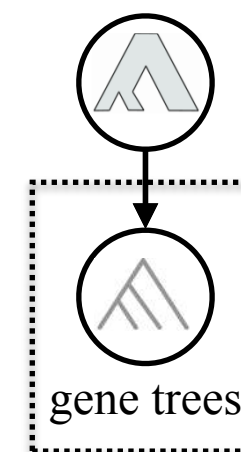
Horizontal gene transfer as information

Transfer events, encoded in the topologies of gene trees can be thought of as “*molecular fossils*” that record the order of speciation events.





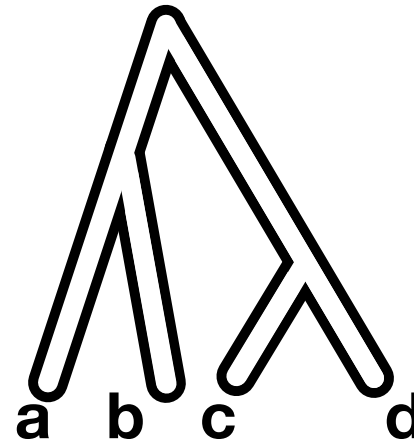
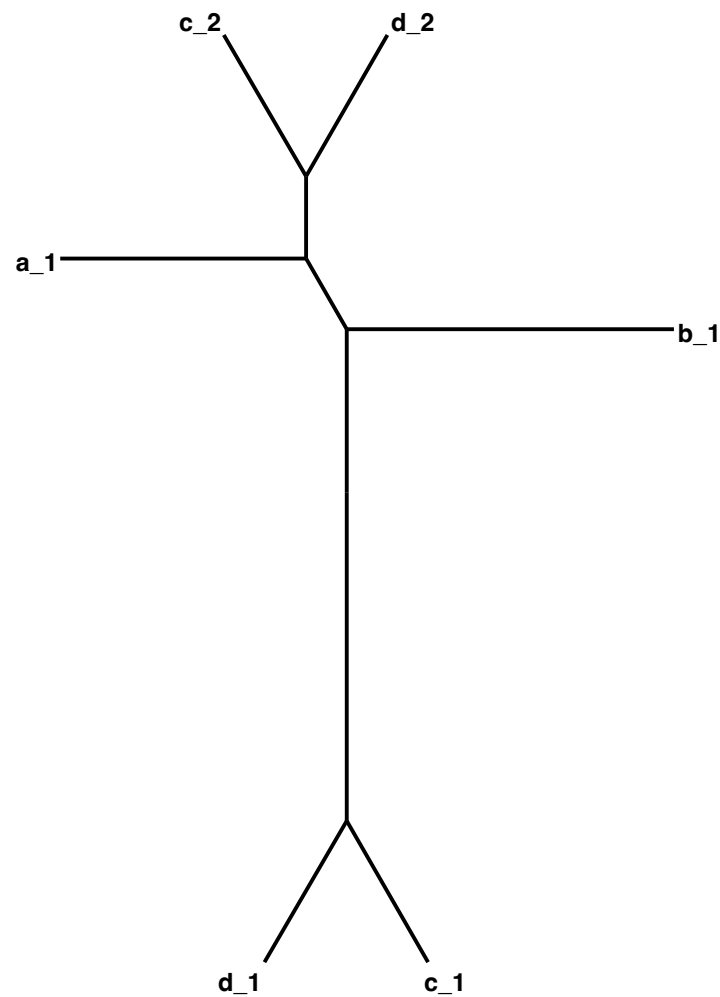
DTL
species tree



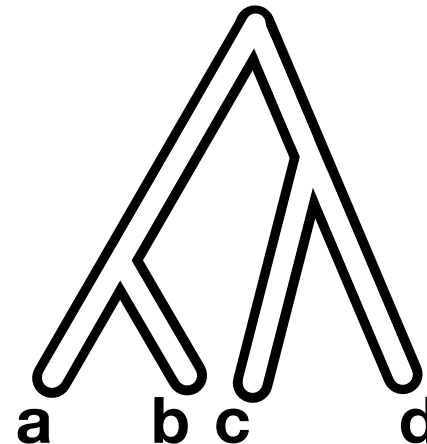
abcd/

```
$ cd ~/workshop_materials/lab_data/abcd
```

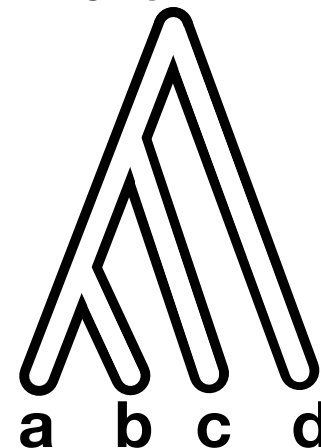
gene tree
g.tree



abcd_S.tree

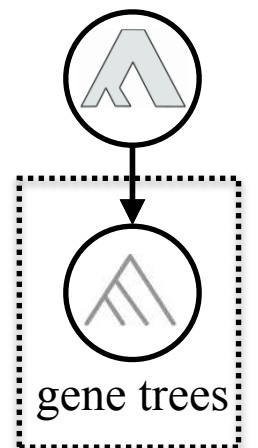


cdab_S.tree



abc-d_S.tree

DTL
species tree



abcd/

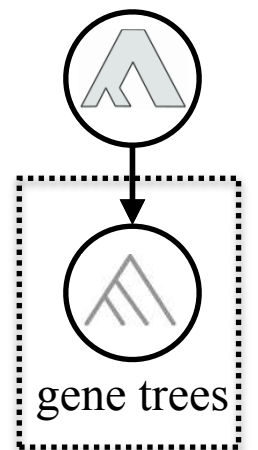
ALEobserve g.tree

ALEml **abcd_S.tree** g.tree.ale sample=10

ALEml **cdab_S.tree** g.tree.ale sample=10

ALEml **abc-d_S.tree** g.tree.ale sample=10

DTL
species tree



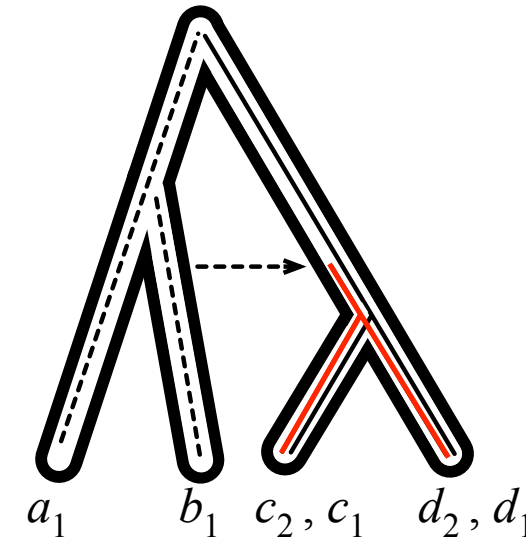
abcd/

abcd_S.tree
ML log-likelihood



```
..
>logl: -5.93738
rate of Duplications Transfers Losses
ML      4.17809e-07  0.372863 1e-10
..
# of Duplications Transfers Losses Speciations
Total 0 1.6 0 3.8

# of Duplications Transfers Losses copies
S_terminal_branch a 0 0 0 0 1
S_terminal_branch b 0 0 0 0 1
S_terminal_branch c 0 0 0 0 2
S_terminal_branch d 0 0 0 0 2
S_internal_branch 1 0 1.2 0 2
S_internal_branch 2 0 0.1 0 1
S_internal_branch 3 0 0.3 0 0.8
abcd_S.tree_g.tree.ale.ml_rec (END)
```



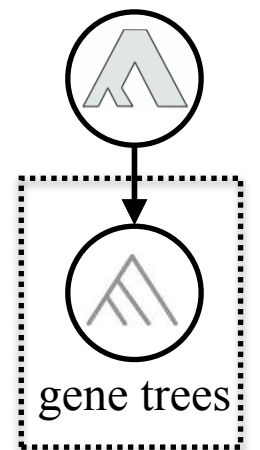
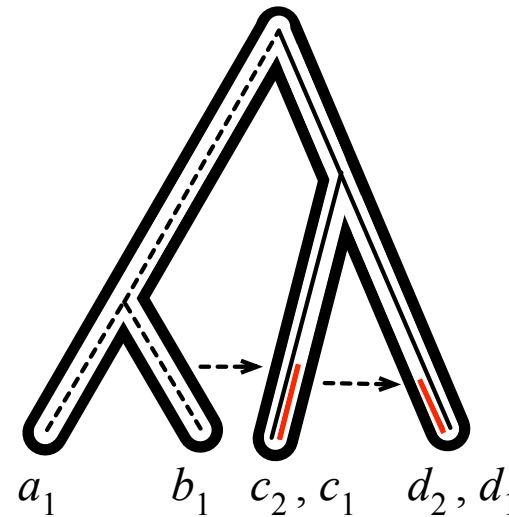
DTL
species tree

cdab_S.tree
ML log-likelihood



```
..
>logl: -9.86434
rate of Duplications Transfers Losses
ML      1e-10  1.05185 1e-10
..
# of Duplications Transfers Losses Speciations
Total 0 3 0 2.3

# of Duplications Transfers Losses copies
S_terminal_branch a 0 0 0.1 0 1
S_terminal_branch b 0 0 0.1 0 1
S_terminal_branch c 0 0 1.2 0 2
S_terminal_branch d 0 0 1.2 0 2
S_internal_branch 1 0 0.2 0 0.9
S_internal_branch 2 0 0.1 0 0.8
S_internal_branch 3 0 0.1 0 0.6
cdab_S.tree_g.tree.ale.ml_rec (END)
```

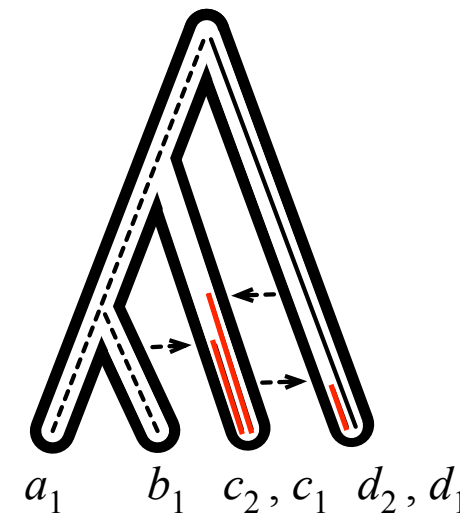


abc-d_S.tree
ML log-likelihood



```
..
>logl: -9.83179
rate of Duplications Transfers Losses
ML      3.74258e-06  1.06978 1e-10
..
# of Duplications Transfers Losses Speciations
Total 0 2.6 0 2.6

# of Duplications Transfers Losses copies
S_terminal_branch a 0 0 0 0 1
S_terminal_branch b 0 0 0 0 1
S_terminal_branch c 0 0 1 0 2
S_terminal_branch d 0 0 1.4 0 2
S_internal_branch 1 0 0 0 1
S_internal_branch 2 0 0.2 0 1
S_internal_branch 3 0 0 0 0.6
abc-d_S.tree_g.tree.ale.ml_rec (END)
```



abcd/

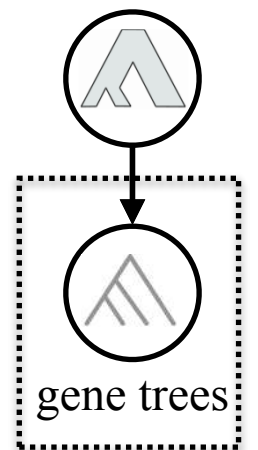
ALEobserve g.tree

ALEml **abcd_S.tree** g.tree.ale sample=10 **delta=0.0 tau=0.01 lambda=0.01**

ALEml **cdab_S.tree** g.tree.ale sample=10 **delta=0.0 tau=0.01 lambda=0.01**

ALEml **abc-d_S.tree** g.tree.ale sample=10 **delta=0.0 tau=0.01 lambda=0.01**

DTL
species tree



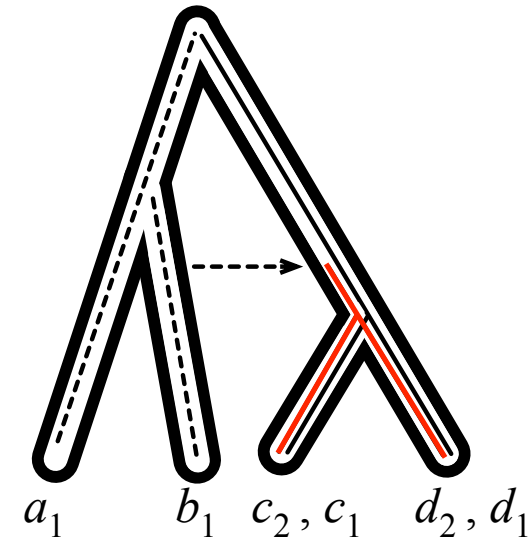
abcd/

log-likelihood →
events →

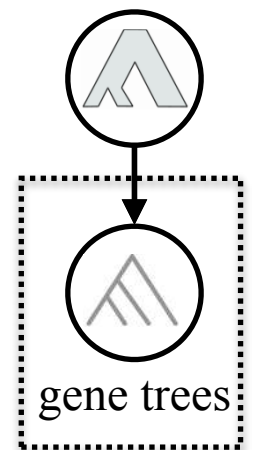
ALEml **abcd_S.tree** g.tree.ale sample=10 delta=0.0 tau=0.01 lambda=0.01

```
..
>logl: -8.61538
..
# of      Duplications      Transfers      Losses      Speciations
Total    0          1          0          4

# of      Duplications      Transfers      Losses      copies
S_terminal_branch  a          0          0          0          1
S_terminal_branch  b          0          0          0          1
S_terminal_branch  c          0          0          0          2
S_terminal_branch  d          0          0          0          2
S_internal_branch  1          0          1          0          2
S_internal_branch  2          0          0          0          1
S_internal_branch  3          0          0          0          1
abcd_S.tree_g.tree.ale.ml_rec (END)
```



DTL
species tree

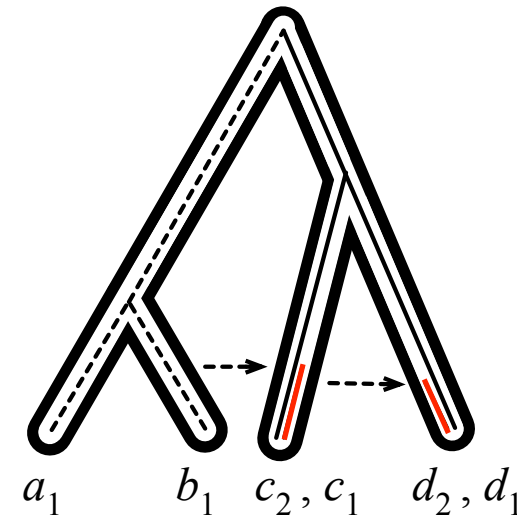


log-likelihood →
events →

ALEml **cdab_S.tree** g.tree.ale sample=10 delta=0.0 tau=0.01 lambda=0.01

```
..
>logl: -16.7168
..
# of      Duplications      Transfers      Losses      Speciations
Total    0          1.9        0.3          3.4

# of      Duplications      Transfers      Losses      copies
S_terminal_branch  a          0          0          0.1          1
S_terminal_branch  b          0          0          0.1          0.2          1
S_terminal_branch  c          0          0          0.8          0          2
S_terminal_branch  d          0          0          0.8          0          2
S_internal_branch  1          0          0          0          1.1
S_internal_branch  2          0          0          0.1          0          1.2
S_internal_branch  3          0          0          0.1          0          1.1
cdab_S.tree_g.tree.ale.ml_rec (END)
```



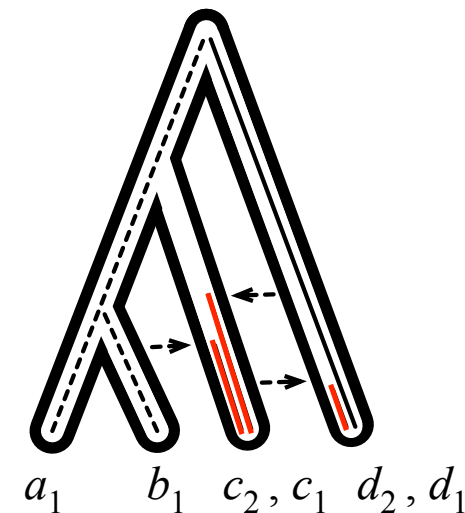
log-likelihood →
events →

ALEml **abc-d_S.tree** g.tree.ale sample=10 delta=0.0 tau=0.01 lambda=0.01

```
..
>logl: -16.8563
rate of      Duplications      Transfers      Losses
ML          3.74258e-06        1.06978 1e-10

..
# of      Duplications      Transfers      Losses      Speciations
Total    0          2          0          3

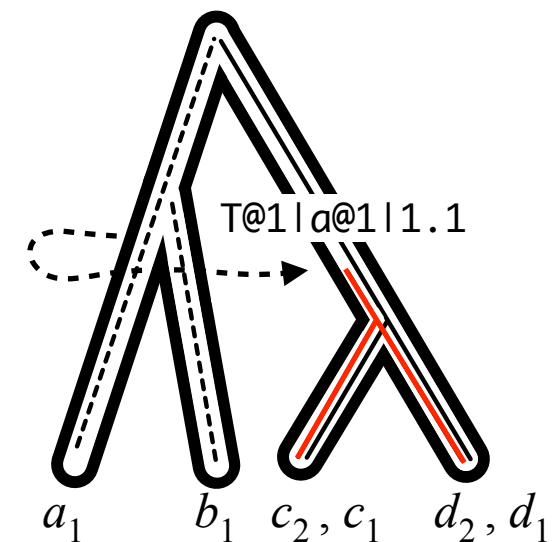
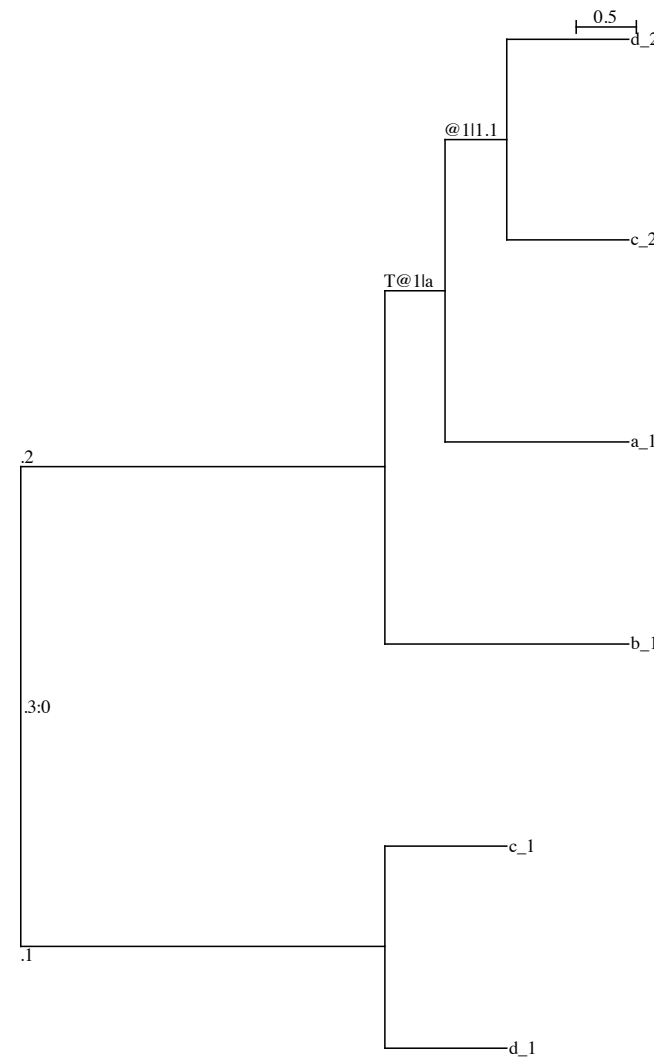
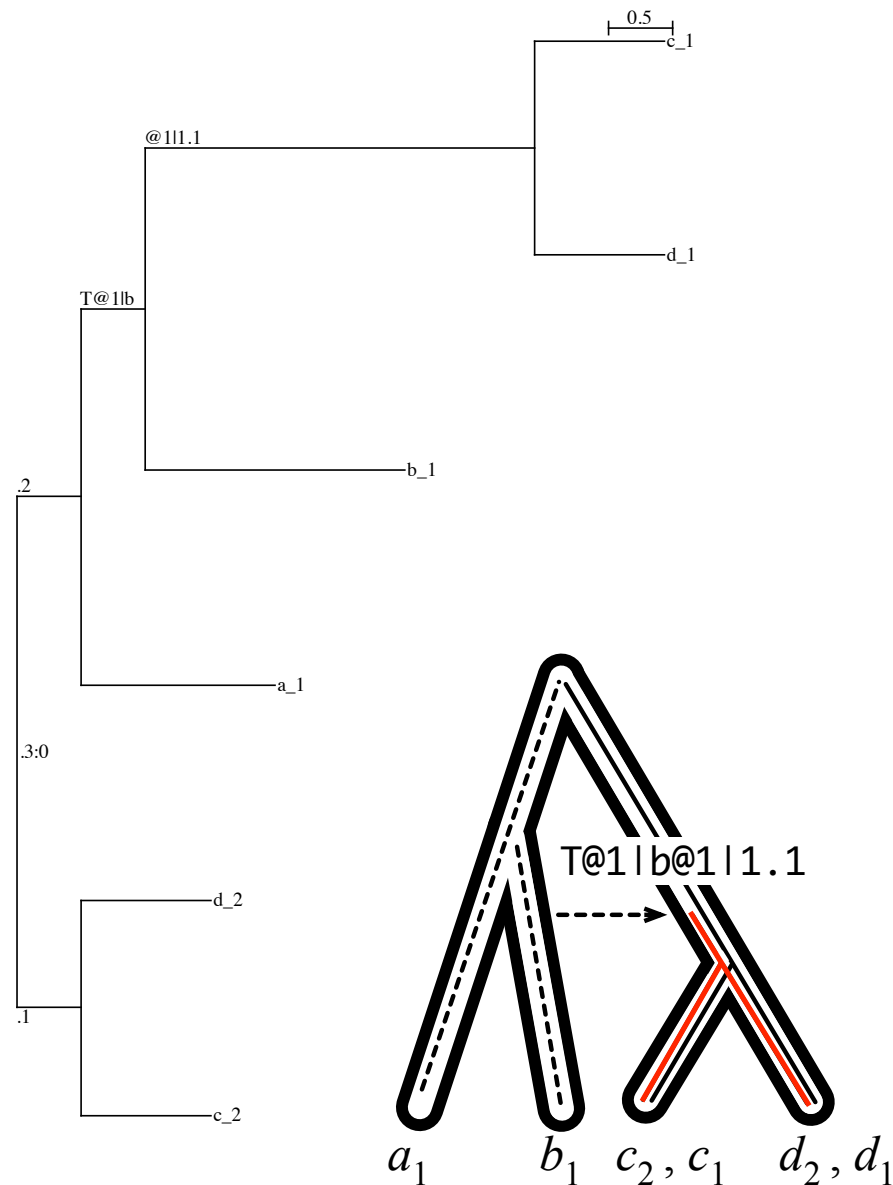
# of      Duplications      Transfers      Losses      copies
S_terminal_branch  a          0          0          0          1
S_terminal_branch  b          0          0          0          1
S_terminal_branch  c          0          0          1          0          2
S_terminal_branch  d          0          0          1          0          2
S_internal_branch  1          0          0          0          1
S_internal_branch  2          0          0          0          1
S_internal_branch  3          0          0          0          1
abc-d_S.tree_g.tree.ale.ml_rec (END)
```



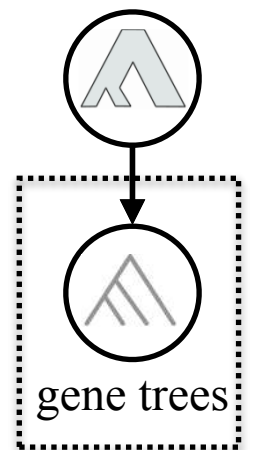
abcd/

ALEml **abcd_S.tree** g.tree.ale sample=10 delta=0.0 tau=0.01 lambda=0.01

```
((d_1:1,c_1:1).1:3,(b_1:2,(a_1:1.5,(c_2:1,d_2:1)@1|1.1:0.5)T@1|a:0.5).2:3).3:0;
((d_1:1,c_1:1).1:3,(b_1:2,(a_1:1.5,(c_2:1,d_2:1)@1|1.1:0.5)T@1|a:0.5).2:3).3:0;
((c_2:1,d_2:1).1:0.5,(a_1:1.5,(b_1:2,(d_1:1,c_1:1)@1|1.1:3)T@1|b:0.5).2:0.5).3:0;
((c_2:1,d_2:1).1:0.5,(a_1:1.5,(b_1:2,(d_1:1,c_1:1)@1|1.1:3)T@1|b:0.5).2:0.5).3:0;
((c_2:1,d_2:1).1:0.5,(a_1:1.5,(b_1:2,(d_1:1,c_1:1)@1|1.1:3)T@1|b:0.5).2:0.5).3:0;
((d_1:1,c_1:1).1:3,(b_1:2,(a_1:1.5,(c_2:1,d_2:1)@1|1.1:0.5)T@1|a:0.5).2:3).3:0;
((d_1:1,c_1:1).1:3,(b_1:2,(a_1:1.5,(c_2:1,d_2:1)@1|1.1:0.5)T@1|a:0.5).2:3).3:0;
((c_2:1,d_2:1).1:0.5,(a_1:1.5,(b_1:2,(d_1:1,c_1:1)@1|1.1:3)T@1|b:0.5).2:0.5).3:0;
((d_1:1,c_1:1).1:3,(b_1:2,(a_1:1.5,(c_2:1,d_2:1)@1|1.1:0.5)T@1|a:0.5).2:3).3:0;
((d_1:1,c_1:1).1:3,(b_1:2,(a_1:1.5,(c_2:1,d_2:1)@1|1.1:0.5)T@1|a:0.5).2:3).3:0;
```



DTL
species tree



view the tree using either
phylo.io, seaview or FigTree!

abcd/

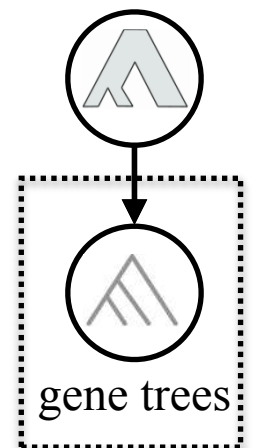
ALEml cdab_S.tree g.tree.ale sample=10 delta=0.0 tau=0.01 lambda=0.01

```
((d_1:1,c_1:1).2:3,(b_1:2,(a_1:1.5,(c_2@0|c:1,d_2@0|d:1)T@0|-1:0.5)T@0|a:0.5).1:3).3:0;  
((a_1.1:1.5,(c_2:1,d_2:1).2:0.5).3:0.5,(b_1.1:2,(d_1:1,c_1:1).2:3)@3|3.3:0.5)T@3|3:0;  
((c_2:1,d_2:1).2:0.5,(a_1:1.5,(b_1:2,(d_1@0|d:1,c_1@0|c:1)T@0|-1:3)T@0|b:0.5).1:0.5).3:0;  
((c_2:1,d_2:1).2:0.5,(a_1:1.5,(b_1:2,(d_1@0|d:1,c_1@0|c:1)T@0|-1:3)T@0|b:0.5).1:0.5).3:0;  
((a_1.1:1.5,(c_2:1,d_2:1).2:0.5).3:0.5,(b_1@0|b:2,(d_1:1,c_1:1)@2|2.2:3)T@2|-1:0.5)T@3|3:0;  
((c_2:1,d_2:1).2:0.5,(a_1:1.5,(b_1:2,(d_1@0|d:1,c_1@0|c:1)T@0|-1:3)T@0|b:0.5).1:0.5).3:0;  
((c_2:1,d_2:1).2:0.5,(a_1:1.5,(b_1:2,(d_1@0|d:1,c_1:1)@0|cT@0|c:3)T@0|b:0.5).1:0.5).3:0;  
((c_2:1,d_2:1).2:0.5,(a_1:1.5,(b_1:2,(d_1@0|d:1,c_1:1)@0|cT@0|c:3)T@0|b:0.5).1:0.5).3:0;  
((c_2:1,d_2:1).2:0.5,(a_1:1.5,(b_1:2,(d_1:1,c_1@0|c:1)@0|dT@0|d:3)T@0|b:0.5).1:0.5).3:0;  
((d_1:1,c_1:1).2:3,(b_1:2,(a_1:1.5,(c_2:1,d_2@0|d:1)@0|cT@0|c:0.5)T@0|a:0.5).1:3).3:0;
```

ALEml cdab_S.tree g.tree.ale sample=10 delta=0.01 tau=0.0 lambda=0.01

```
Total    0      2      0      3  
((a_1.1:1.5,(c_2:1,d_2:1).2:0.5).3:0.5,(b_1.1:2,(d_1:1,c_1:1).2:3).3:0.5)D@3|1.45|3:0;  
((a_1.1:1.5,(c_2:1,d_2:1).2:0.5).3:0.5,(b_1.1:2,(d_1:1,c_1:1).2:3).3:0.5)D@3|1.3|3:0;  
((a_1.1:1.5,(c_2:1,d_2:1).2:0.5).3:0.5,(b_1.1:2,(d_1:1,c_1:1).2:3).3:0.5)D@3|1.2|3:0;  
((a_1.1:1.5,(c_2:1,d_2:1).2:0.5).3:0.5,(b_1.1:2,(d_1:1,c_1:1).2:3).3:0.5)D@3|1.4|3:0;  
((a_1.1:1.5,(c_2:1,d_2:1).2:0.5).3:0.5,(b_1.1:2,(d_1:1,c_1:1).2:3).3:0.5)D@3|1.1|3:0;  
((a_1.1:1.5,(c_2:1,d_2:1).2:0.5).3:0.5,(b_1.1:2,(d_1:1,c_1:1).2:3).3:0.5)D@3|1.45|3:0;  
((a_1.1:1.5,(c_2:1,d_2:1).2:0.5).3:0.5,(b_1.1:2,(d_1:1,c_1:1).2:3).3:0.5)D@3|1.2|3:0;  
((a_1.1:1.5,(c_2:1,d_2:1).2:0.5).3:0.5,(b_1.1:2,(d_1:1,c_1:1).2:3).3:0.5)D@3|1.1|3:0;  
((a_1.1:1.5,(c_2:1,d_2:1).2:0.5).3:0.5,(b_1.1:2,(d_1:1,c_1:1).2:3).3:0.5)D@3|1.35|3:0;  
((a_1.1:1.5,(c_2:1,d_2:1).2:0.5).3:0.5,(b_1.1:2,(d_1:1,c_1:1).2:3).3:0.5)D@3|1.3|3:0;  
# of      Duplications      Transfers      Losses      Speciations  
Total    1          0          2          6
```

DTL
species tree



view the trees using either
phylo.io, seaview or FigTree!

abcd/

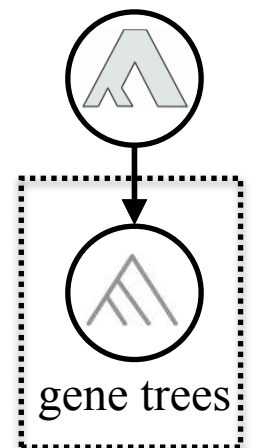
ALEml cdab_S.tree g.tree.ale sample=10 delta=0.0 tau=0.01 lambda=0.01

```
((d_1:1,c_1:1).2:3,(b_1:2,(a_1:1.5,(c_2@0|c:1,d_2@0|d:1)T@0|-1:0.5)T@0|a:0.5).1:3).3:0;
((a_1:1:1.5,(c_2:1,d_2:1).2:0.5).3:0.5,(b_1:1:2,(d_1:1,c_1:1).2:3)@3|3.3:0.5)T@3|3:0;
((c_2:1,d_2:1).2:0.5,(a_1:1.5,(b_1:2,(d_1@0|d:1,c_1@0|c:1)T@0|-1:3)T@0|b:0.5).1:0.5).3:0;
((c_2:1,d_2:1).2:0.5,(a_1:1.5,(b_1:2,(d_1@0|d:1,c_1@0|c:1)T@0|-1:3)T@0|b:0.5).1:0.5).3:0;
((a_1:1:1.5,(c_2:1,d_2:1).2:0.5).3:0.5,(b_1@0|b:2,(d_1:1,c_1:1)@2|2.2:3)T@2|-1:0.5)T@3|3:0;
((c_2:1,d_2:1).2:0.5,(a_1:1.5,(b_1:2,(d_1@0|d:1,c_1@0|c:1)T@0|-1:3)T@0|b:0.5).1:0.5).3:0;
((c_2:1,d_2:1).2:0.5,(a_1:1.5,(b_1:2,(d_1@0|d:1,c_1:1)@0|cT@0|c:3)T@0|b:0.5).1:0.5).3:0;
((c_2:1,d_2:1).2:0.5,(a_1:1.5,(b_1:2,(d_1@0|d:1,c_1:1)@0|cT@0|c:3)T@0|b:0.5).1:0.5).3:0;
((c_2:1,d_2:1).2:0.5,(a_1:1.5,(b_1:2,(d_1:1,c_1@0|c:1)@0|dT@0|d:3)T@0|b:0.5).1:0.5).3:0;
((d_1:1,c_1:1).2:3,(b_1:2,(a_1:1.5,(c_2:1,d_2@0|d:1)@0|cT@0|c:0.5)T@0|a:0.5).1:3).3:0;
```

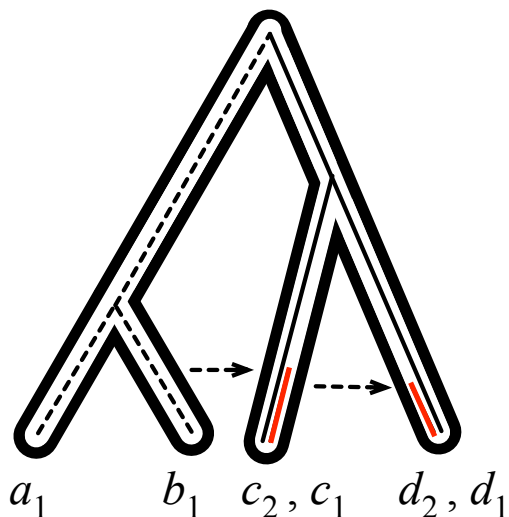
ALEml cdab_S.tree g.tree.ale sample=10 delta=0.01 tau=0.0 lambda=0.01

```
Total    0      2      0      3
((a_1:1:1.5,(c_2:1,d_2:1).2:0.5).3:0.5,(b_1:1:2,(d_1:1,c_1:1).2:3).3:0.5)D@3|1.45|3:0;
((a_1:1:1.5,(c_2:1,d_2:1).2:0.5).3:0.5,(b_1:1:2,(d_1:1,c_1:1).2:3).3:0.5)D@3|1.3|3:0;
((a_1:1:1.5,(c_2:1,d_2:1).2:0.5).3:0.5,(b_1:1:2,(d_1:1,c_1:1).2:3).3:0.5)D@3|1.2|3:0;
((a_1:1:1.5,(c_2:1,d_2:1).2:0.5).3:0.5,(b_1:1:2,(d_1:1,c_1:1).2:3).3:0.5)D@3|1.4|3:0;
((a_1:1:1.5,(c_2:1,d_2:1).2:0.5).3:0.5,(b_1:1:2,(d_1:1,c_1:1).2:3).3:0.5)D@3|1.1|3:0;
((a_1:1:1.5,(c_2:1,d_2:1).2:0.5).3:0.5,(b_1:1:2,(d_1:1,c_1:1).2:3).3:0.5)D@3|1.45|3:0;
((a_1:1:1.5,(c_2:1,d_2:1).2:0.5).3:0.5,(b_1:1:2,(d_1:1,c_1:1).2:3).3:0.5)D@3|1.2|3:0;
((a_1:1:1.5,(c_2:1,d_2:1).2:0.5).3:0.5,(b_1:1:2,(d_1:1,c_1:1).2:3).3:0.5)D@3|1.1|3:0;
((a_1:1:1.5,(c_2:1,d_2:1).2:0.5).3:0.5,(b_1:1:2,(d_1:1,c_1:1).2:3).3:0.5)D@3|1.35|3:0;
((a_1:1:1.5,(c_2:1,d_2:1).2:0.5).3:0.5,(b_1:1:2,(d_1:1,c_1:1).2:3).3:0.5)D@3|1.3|3:0;
# of      Duplications  Transfers      Losses  Speciations
Total    1          0          2          6
```

DTL
species tree



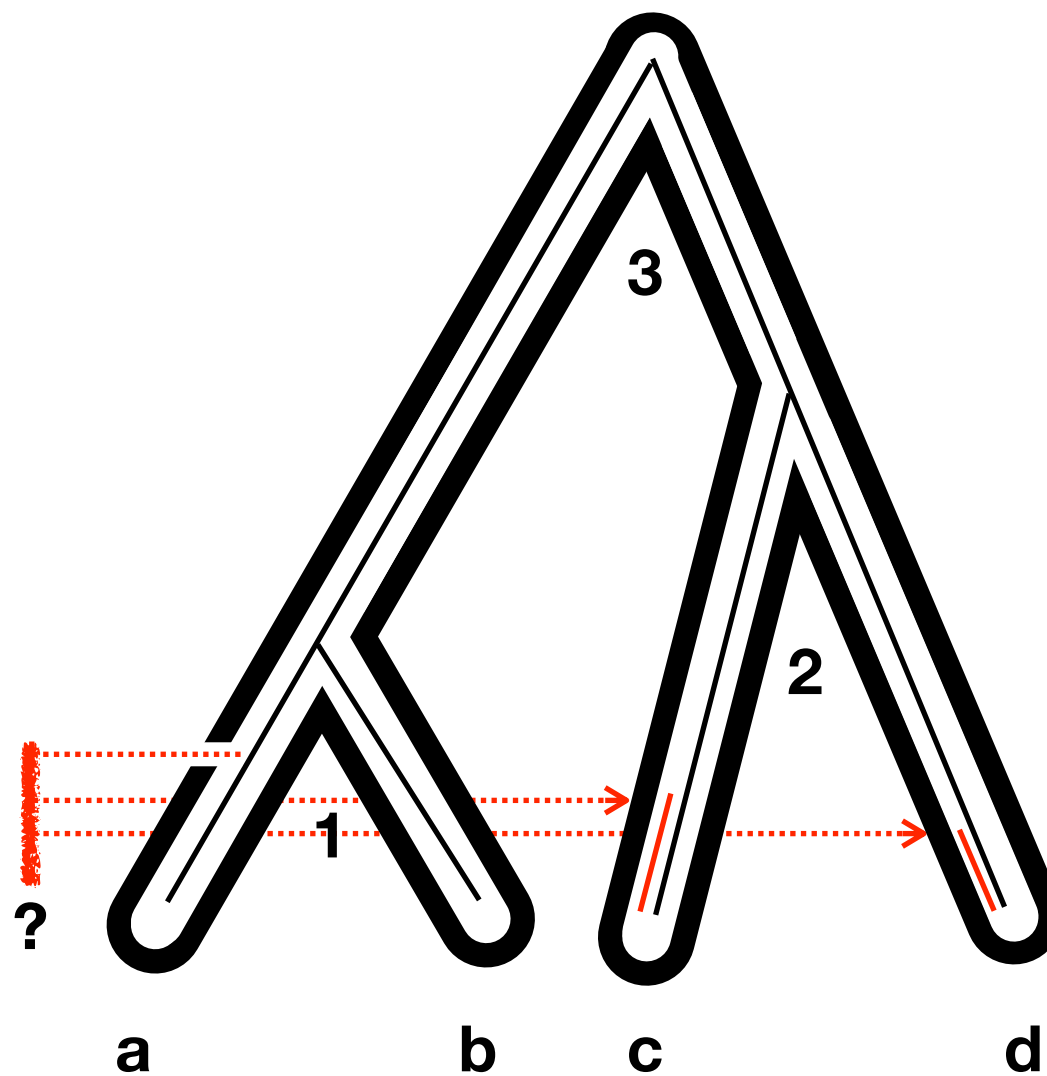
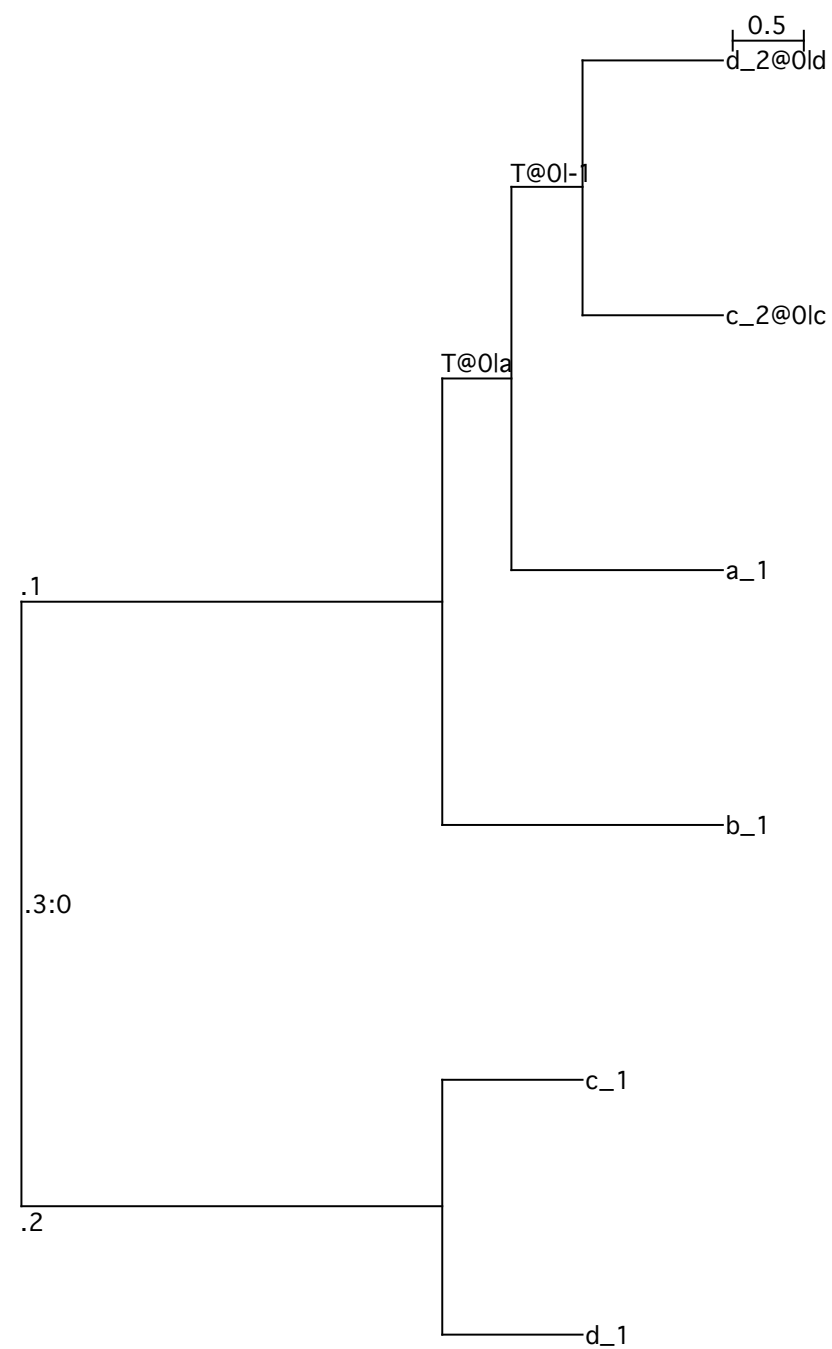
ALEml cdab_S.tree g.tree.ale sample=10 delta=0.0 tau=0.01 lambda=0.0



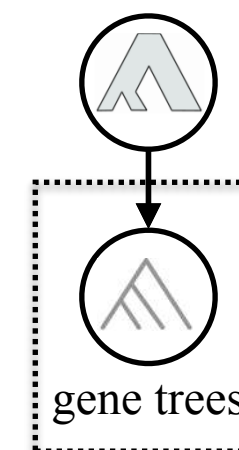
```
((d_1:1,c_1:1).2:3,(b_1:2,(a_1:1.5,(c_2:1,d_2@0|d:1)@0|cT@0|c:0.5)T@0|a:0.5).1:3).3:0;
((c_2:1,d_2:1).2:0.5,(a_1:1.5,(b_1:2,(d_1@0|d:1,c_1@0|c:1)T@0|-1:3)T@0|b:0.5).1:0.5).3:0;
((d_1:1,c_1:1).2:3,(b_1:2,(a_1:1.5,(c_2@0|c:1,d_2@0|d:1)T@0|-1:0.5)T@0|a:0.5).1:3).3:0;
((c_2:1,d_2:1).2:0.5,(a_1:1.5,(b_1:2,(d_1@0|d:1,c_1:1)@0|cT@0|c:3)T@0|b:0.5).1:0.5).3:0;
((d_1:1,c_1:1).2:3,(b_1:2,(a_1:1.5,(c_2@0|c:1,d_2:1)@0|dT@0|d:0.5)T@0|a:0.5).1:3).3:0;
((d_1:1,c_1:1).2:3,(b_1:2,(a_1:1.5,(c_2@0|c:1,d_2:1)@0|dT@0|d:0.5)T@0|a:0.5).1:3).3:0;
((d_1:1,c_1:1).2:3,(b_1:2,(a_1:1.5,(c_2@0|c:1,d_2@0|d:1)T@0|-1:0.5)T@0|a:0.5).1:3).3:0;
((c_2:1,d_2:1).2:0.5,(a_1:1.5,(b_1:2,(d_1@0|d:1,c_1@0|c:1)T@0|-1:3)T@0|b:0.5).1:0.5).3:0;
((c_2:1,d_2:1).2:0.5,(a_1:1.5,(b_1:2,(d_1:1,c_1@0|c:1)@0|dT@0|d:3)T@0|b:0.5).1:0.5).3:0;
((d_1:1,c_1:1).2:3,(b_1:2,(a_1:1.5,(c_2@0|c:1,d_2:1)@0|dT@0|d:0.5)T@0|a:0.5).1:3).3:0;
# of      Duplications  Transfers      Losses  Speciations
```

?
T@0|-1

abcd/



DTL
species tree



view the tree using either
phylo.io, seaview or FigTree!

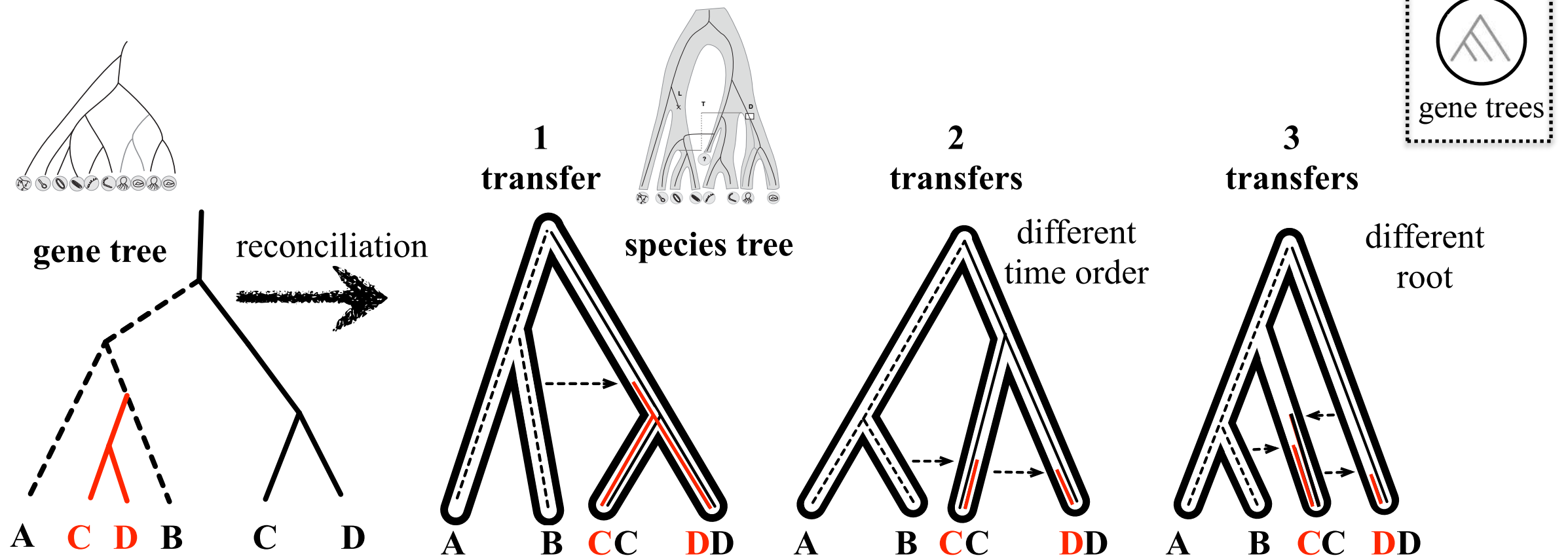
abcd/

ALEobserve g.tree

ALEml **abcd_S.tree** g.tree.ale sample=10

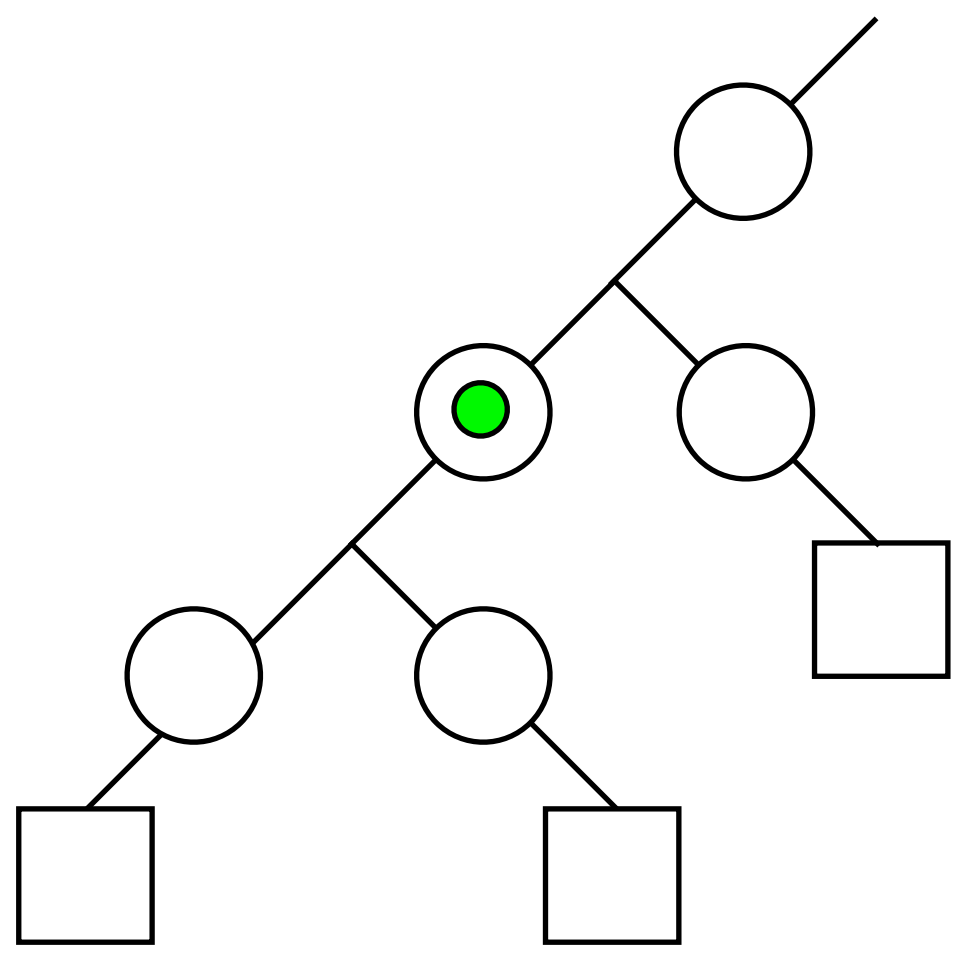
ALEml **cdab_S.tree** g.tree.ale sample=10

ALEml **abc-d_S.tree** g.tree.ale sample=10

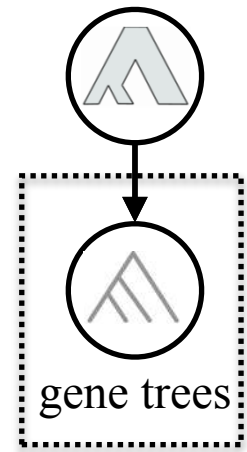


abcd/

“undated”
DTL

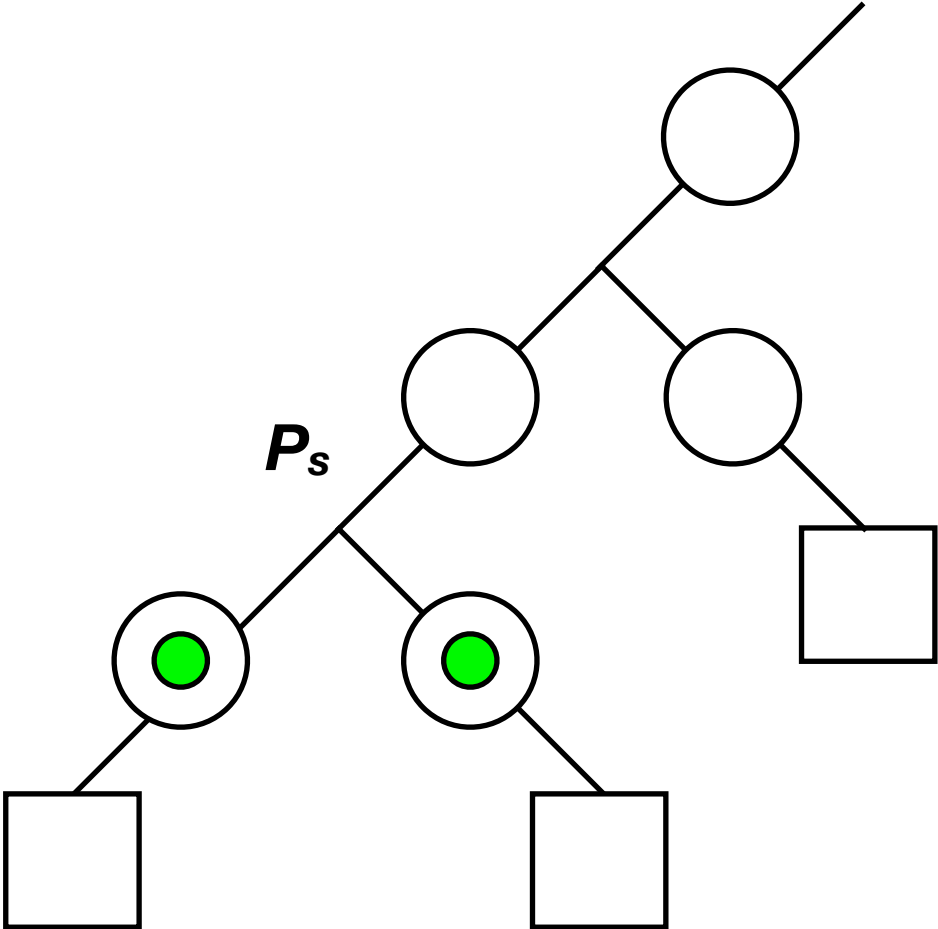


“undated”
DTL
species tree

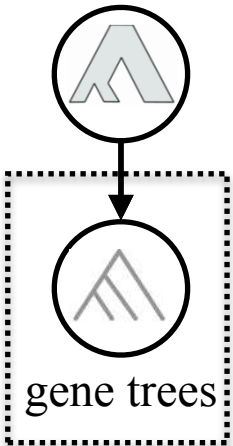


abcd/

“undated” DTL

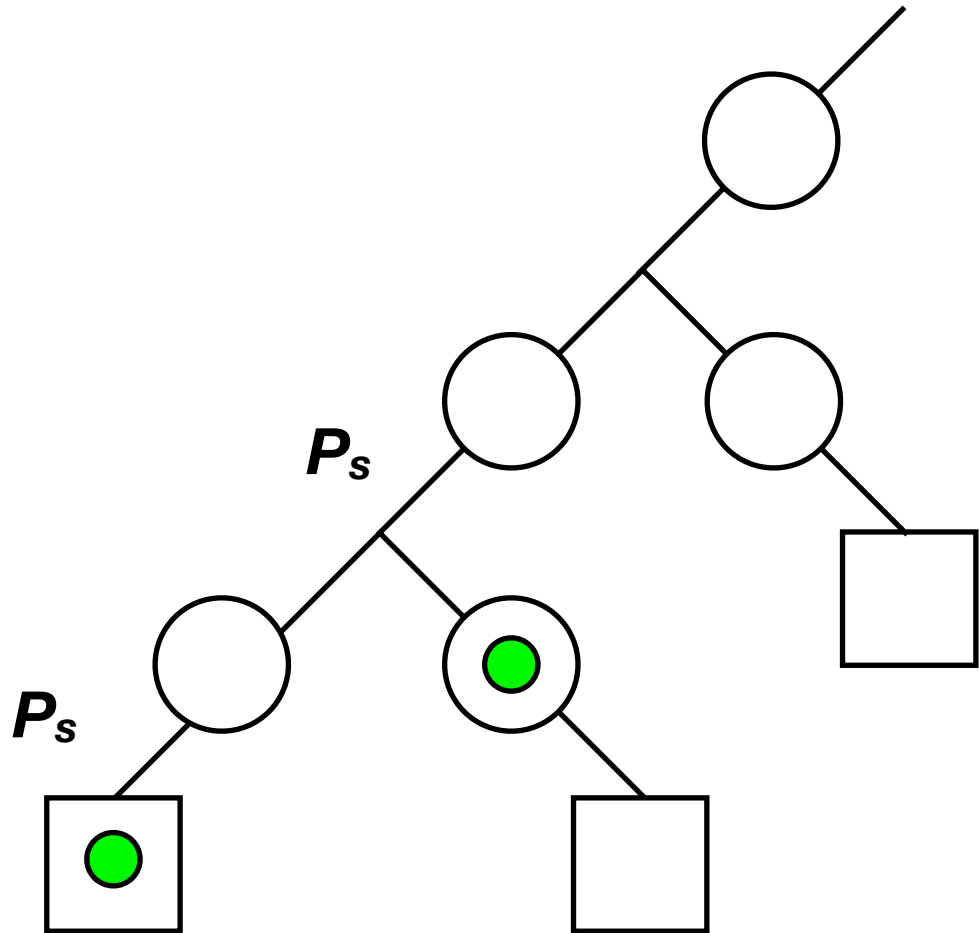


“undated”
DTL
species tree

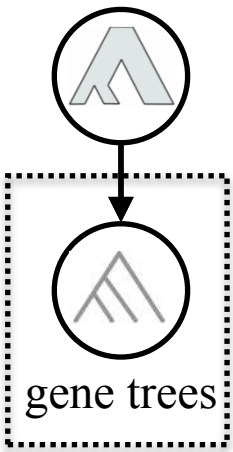


abcd/

“undated”
DTL

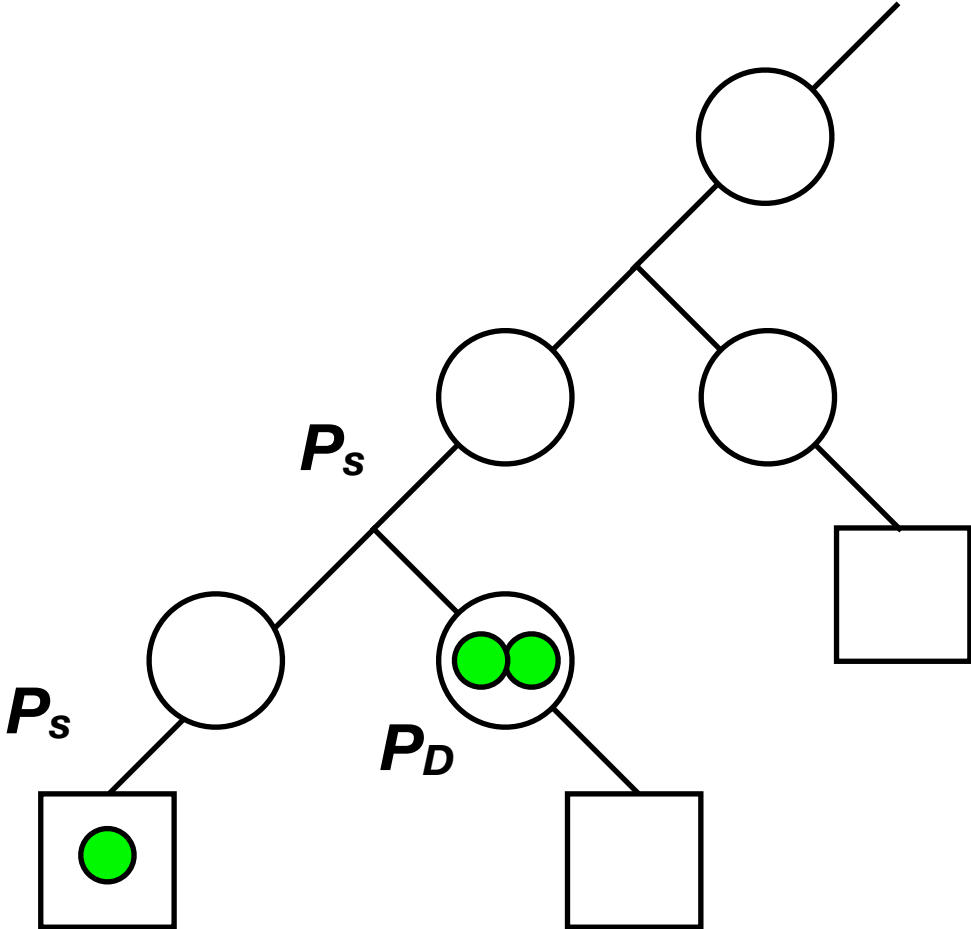


“undated”
DTL
species tree

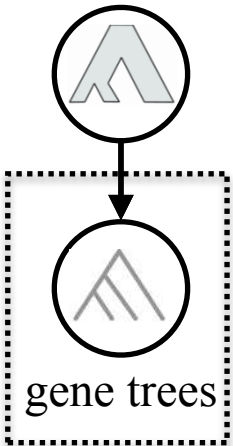


abcd/

“undated” DTL

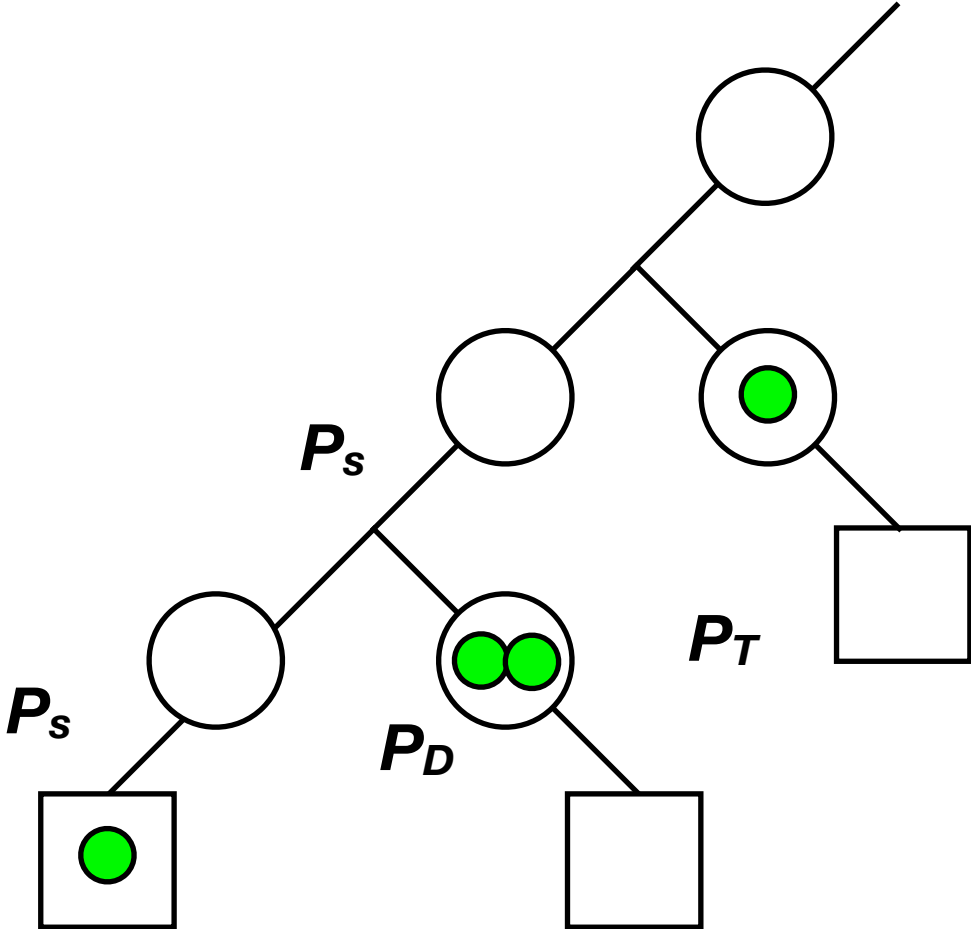


“undated”
DTL
species tree

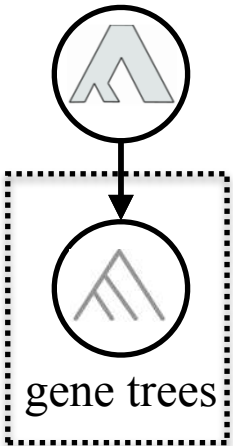


abcd/

“undated” DTL

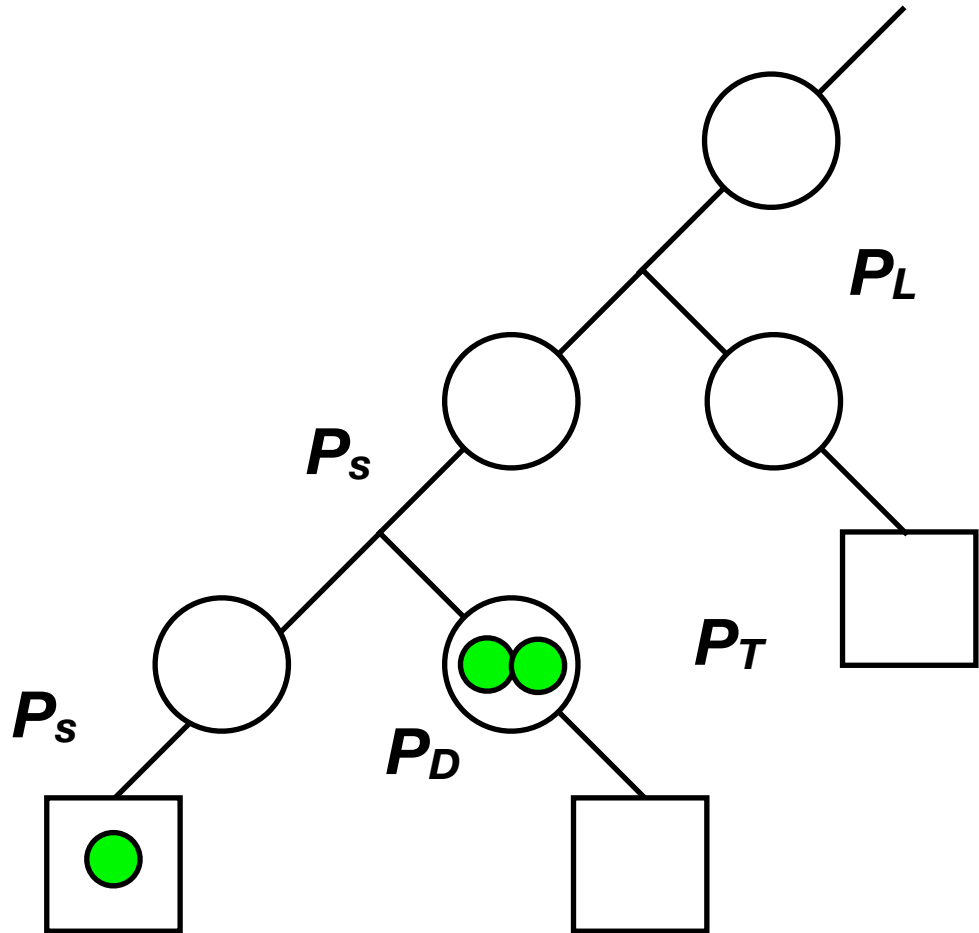


“undated”
DTL
species tree

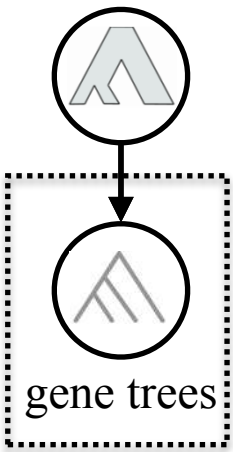


abcd/

“undated”
DTL



“undated”
DTL
species tree



abcd/

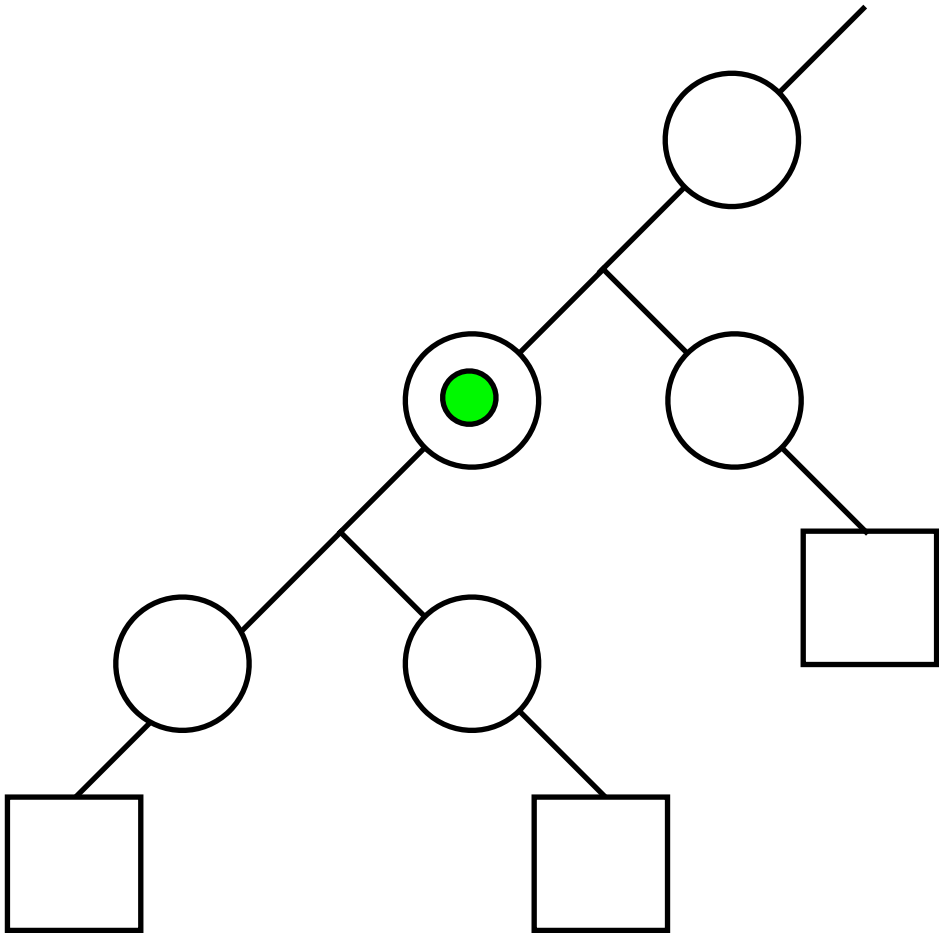
“undated”
DTL

What do you see that is different compared to ALEml?

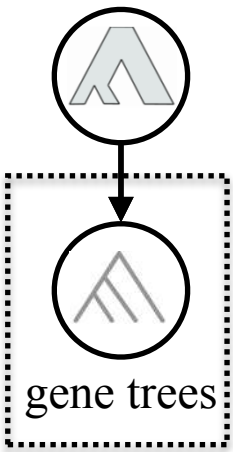
ALEml_undated **abcd_S.tree** g.tree.ale sample=10

ALEml_undated **cdab_S.tree** g.tree.ale sample=10

ALEml_undated **abc-d_S.tree** g.tree.ale sample=10



“undated”
DTL
species tree



abcd/

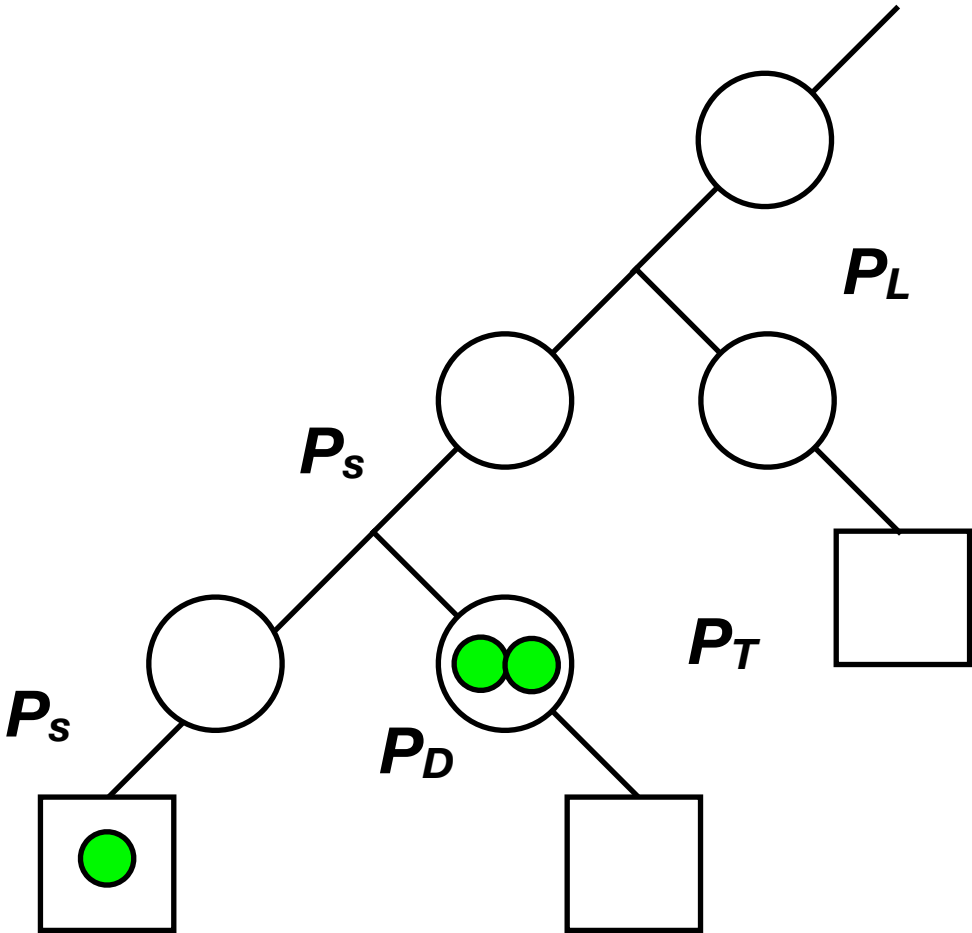
“undated”
DTL

What do you see that is different compared to ALEml?

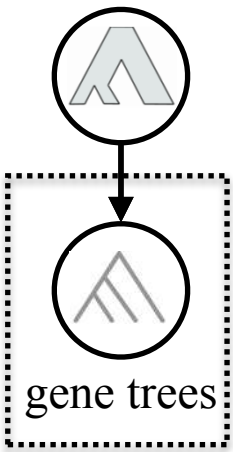
```
ALEml_undated abcd_S.tree g.tree.ale sample=10
```

```
ALEml_undated cdab_S.tree g.tree.ale sample=10
```

```
ALEml_undated abc-d_S.tree g.tree.ale sample=10
```



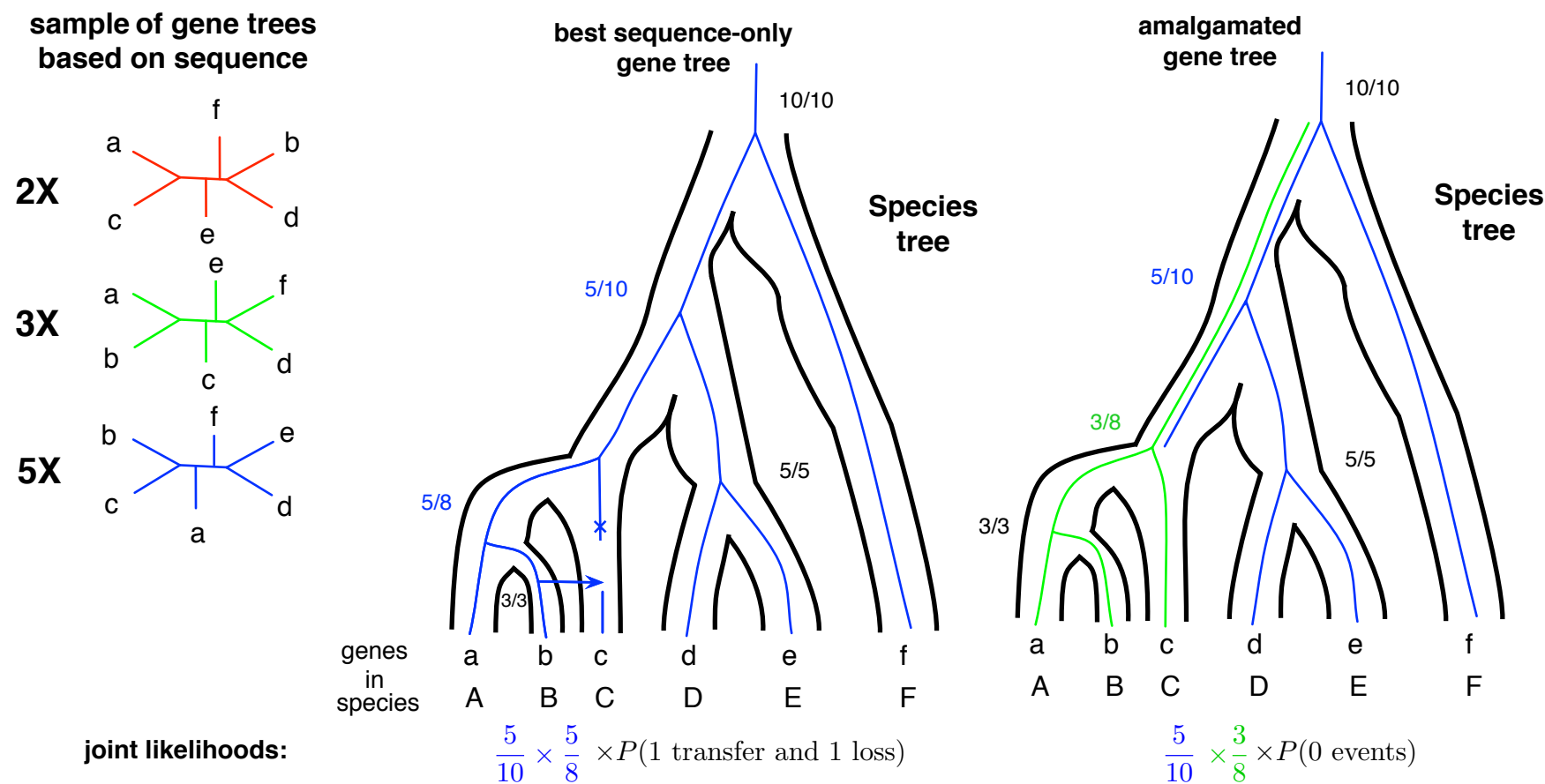
“undated”
DTL
species tree



abcdef/

Efficiently exploring the space of reconciled gene trees

Based on a sample of trees conditional clade probabilities can be used to estimate posterior probability of any gene tree that can be amalgamated. This is usually a very large number of trees (e.g. for 10^4 samples 10^{12} trees, but up to 10^{40}). *The dynamic programming used in gene tree-species tree reconciliation can be extended to approximate the joint likelihood efficiently for a very large set of gene trees.*

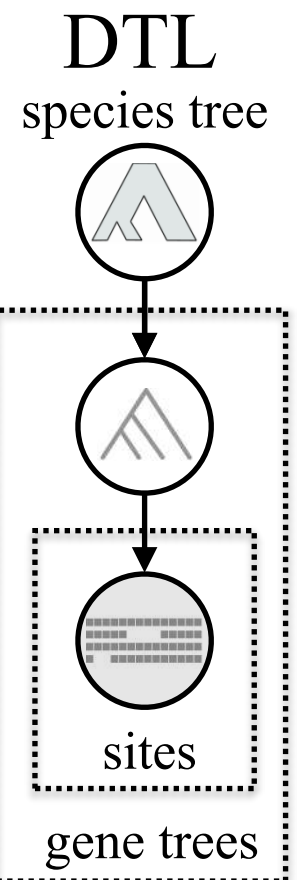


implemented in ALE:

<http://github.com/ssolo/ALE>

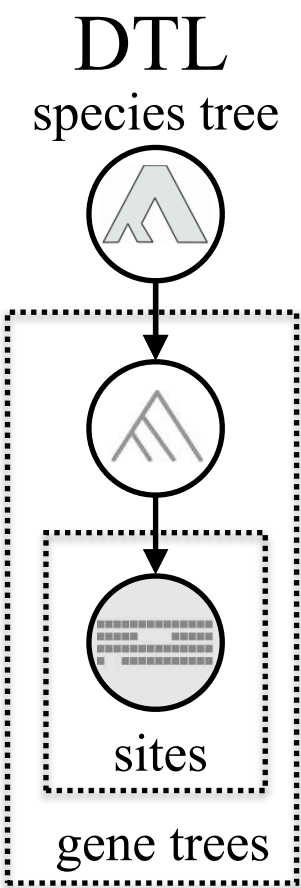
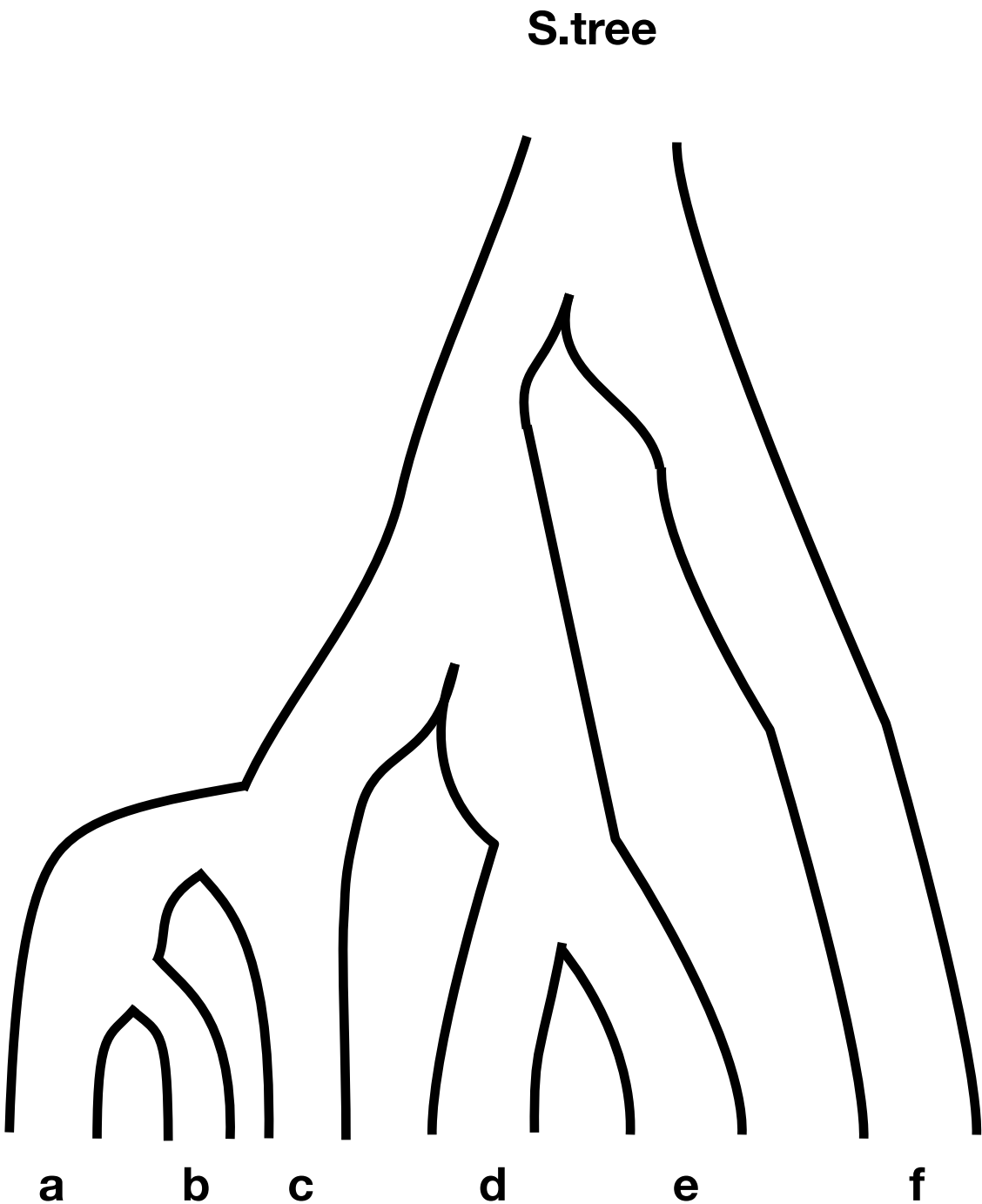
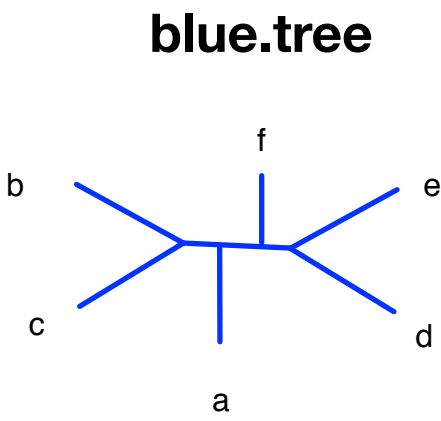
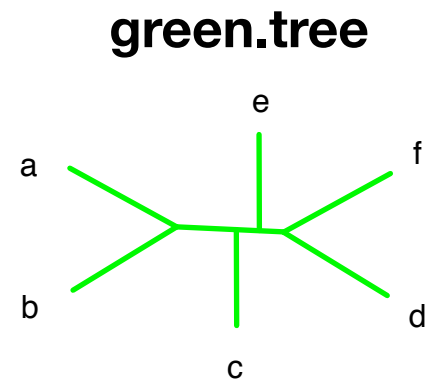
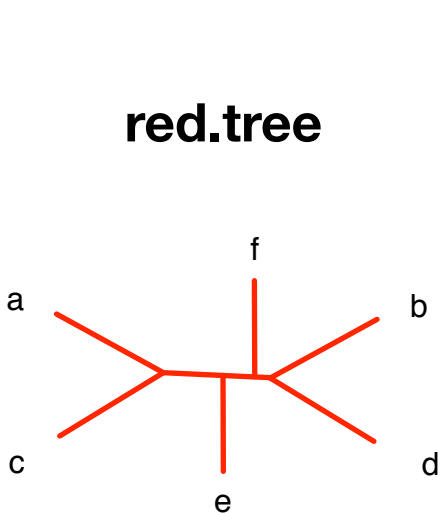
Szöllősi, Tannier, Lartillot & Daubin *Systematic Biology* (2013)
Lateral Gene Transfer from the Dead

Szöllősi, Rosikiewicz, Boussau, Tannier & Daubin *Systematic Biology* (2013)
Efficient exploration of the space of reconciled gene trees



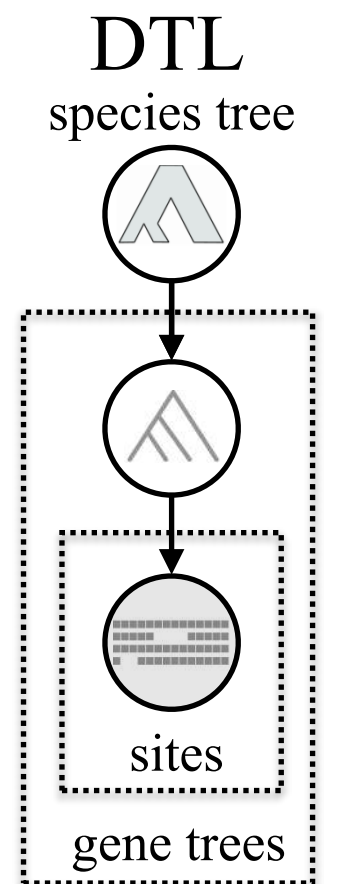
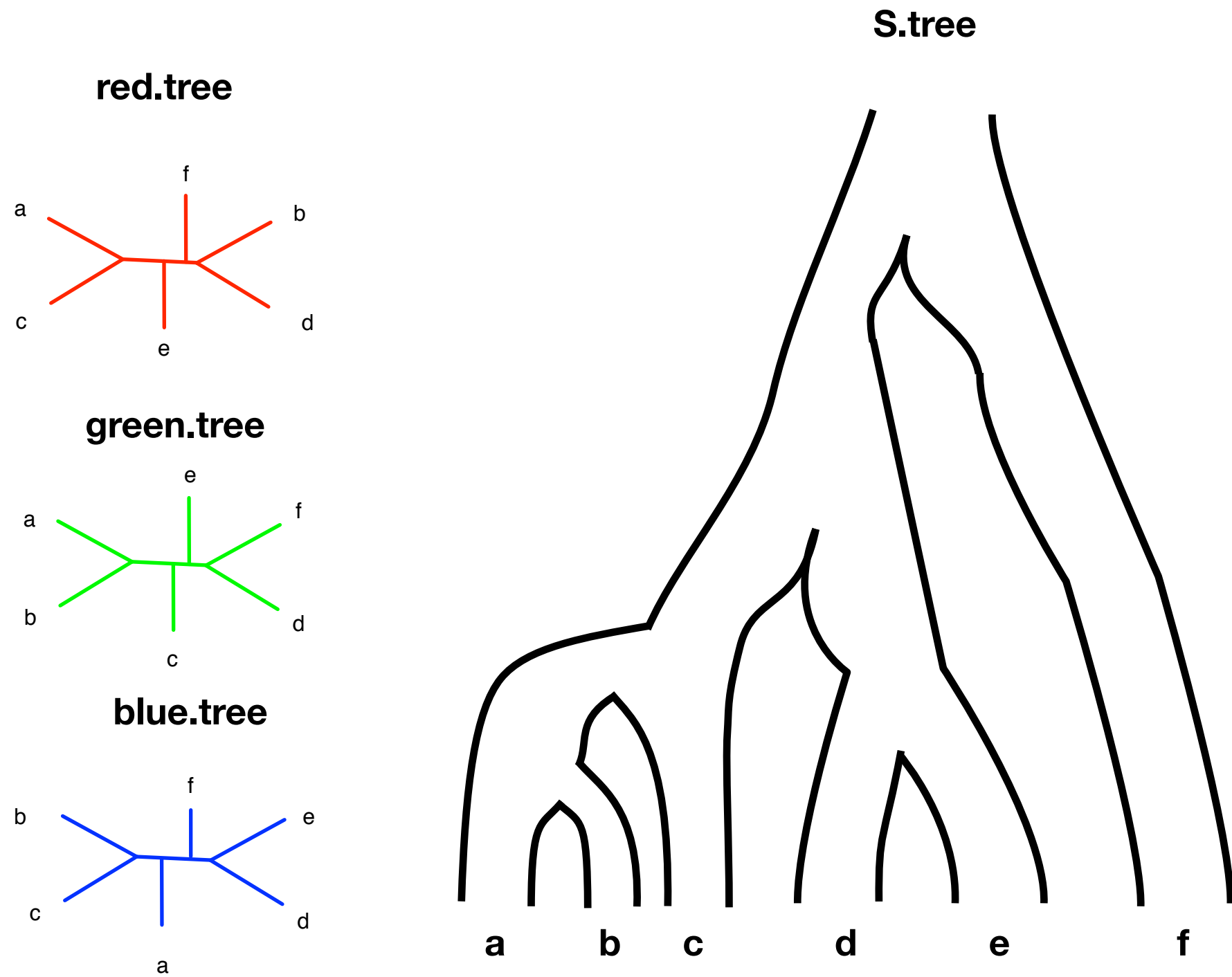
abcdef/

```
$ cd ~/workshop_materials/lab_data/abcdef
```



abcdef/

Use ALEml (and/or ALEml_undated) on each gene tree to estimate **ML rates**
and sample reconciliations with the specie tree



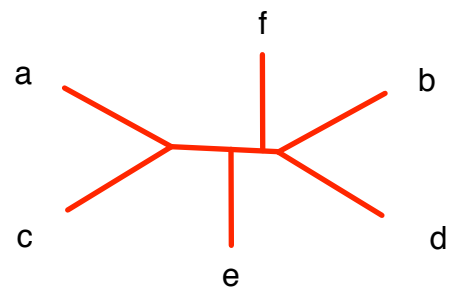
abcdef/

ALEml S.tree **blue.tree.ale**

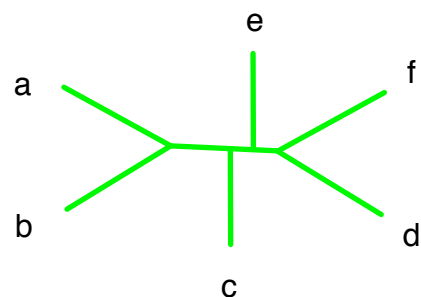
ALEml S.tree **green.tree.ale**

ALEml S.tree **red.tree.ale**

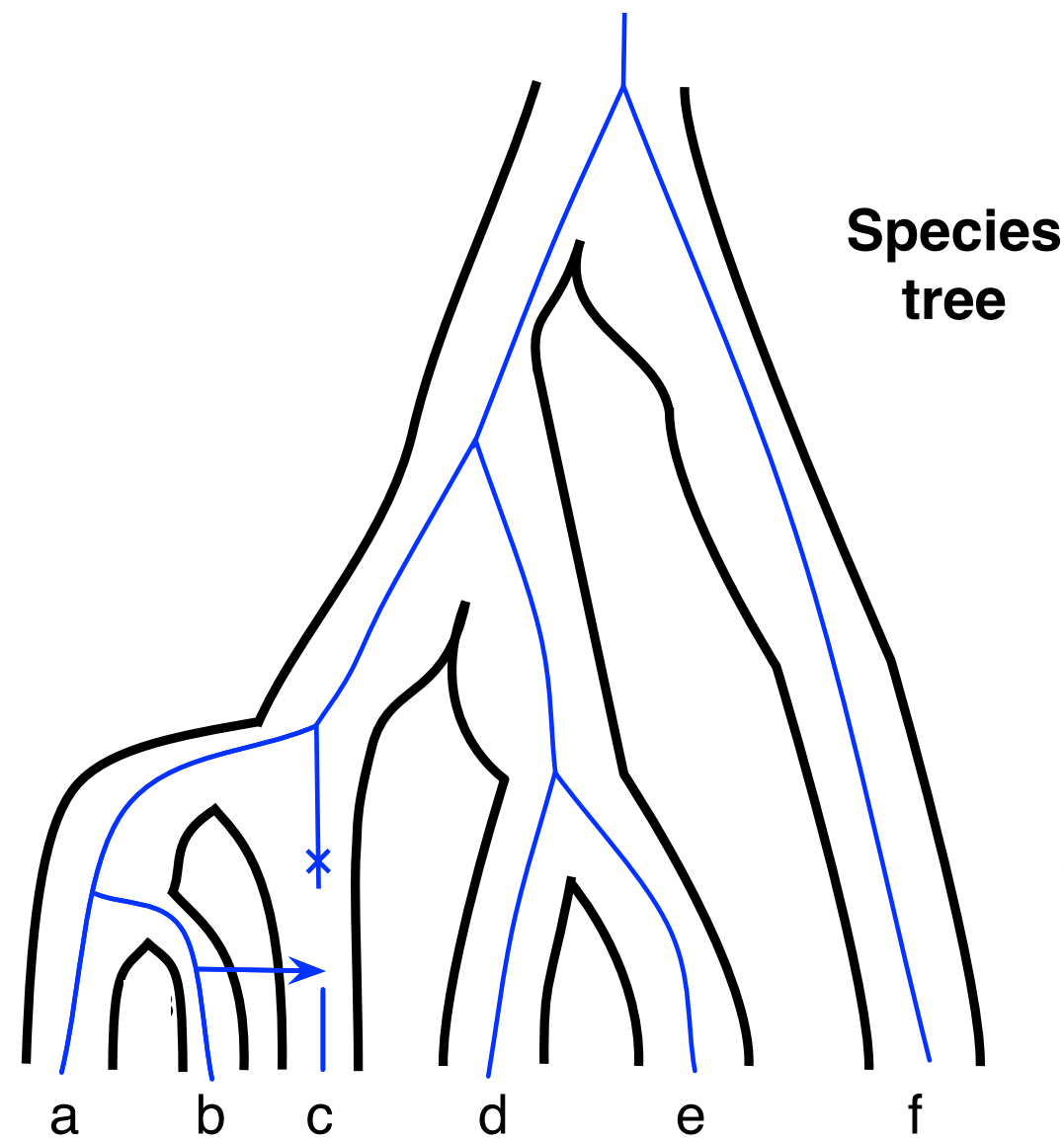
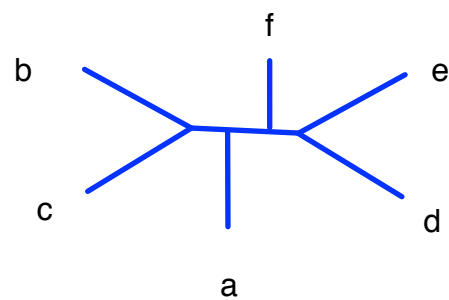
red.tree



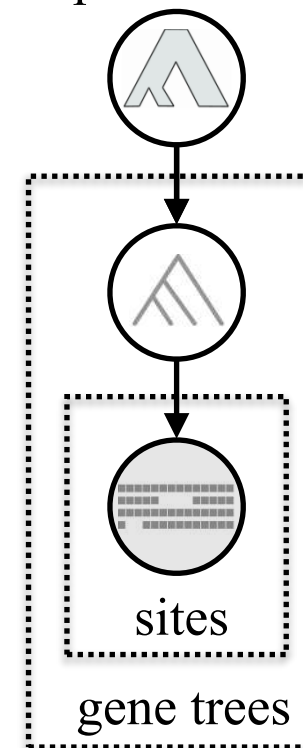
green.tree



blue.tree



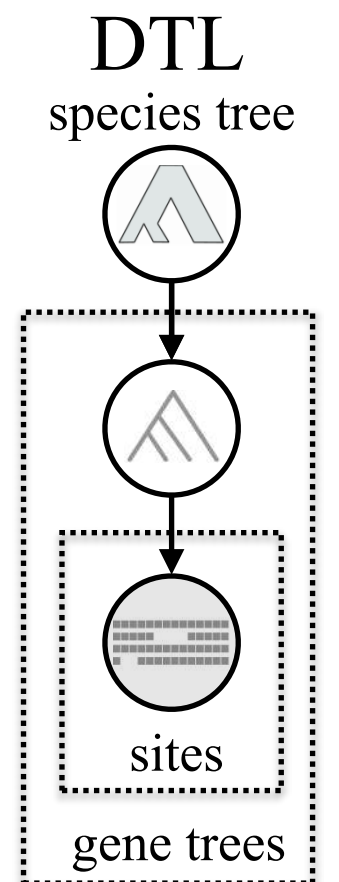
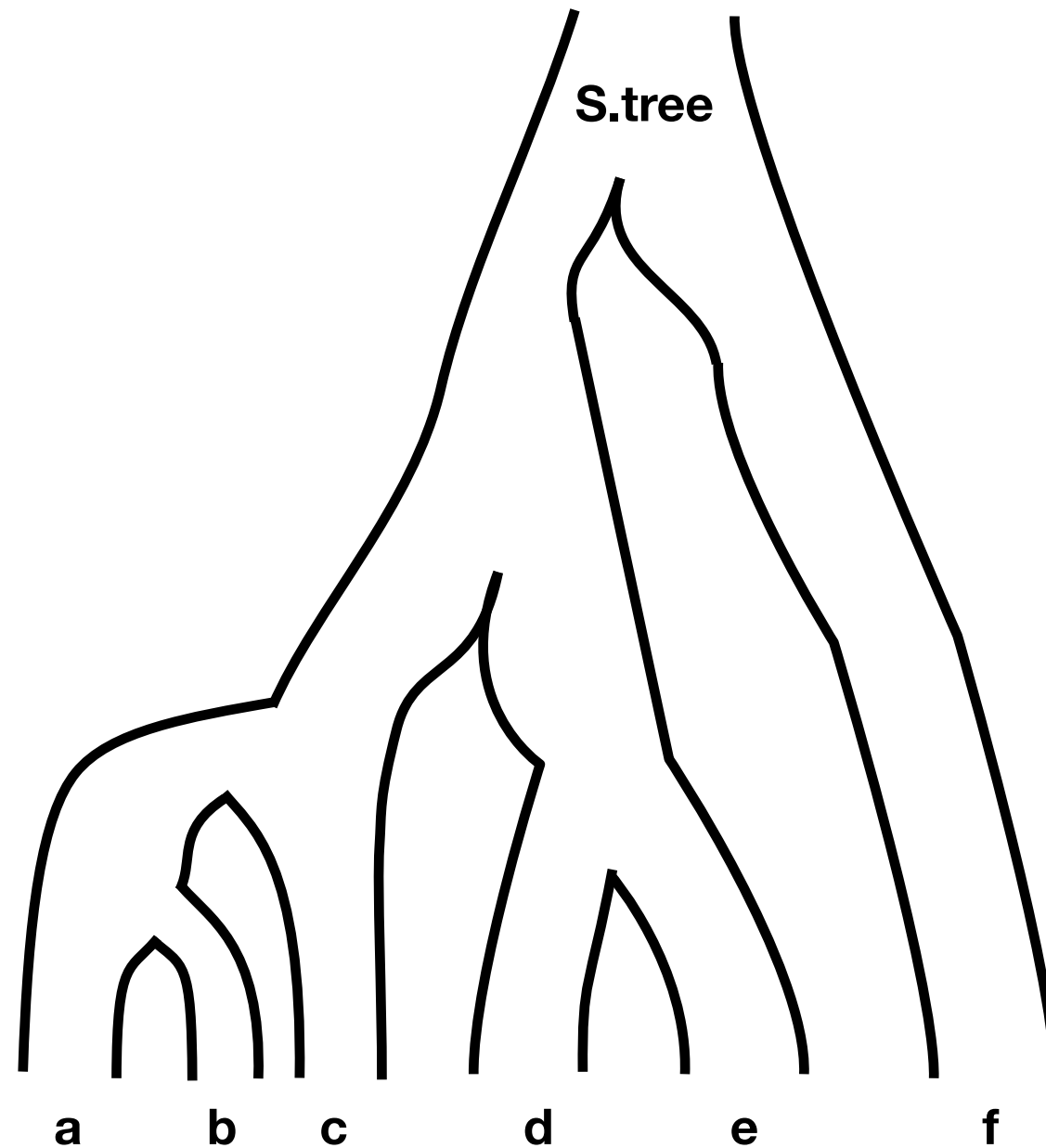
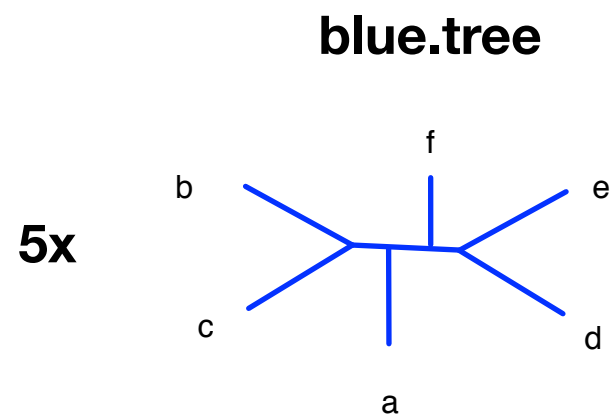
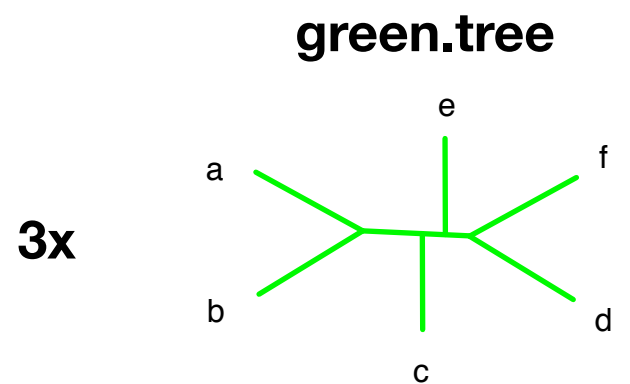
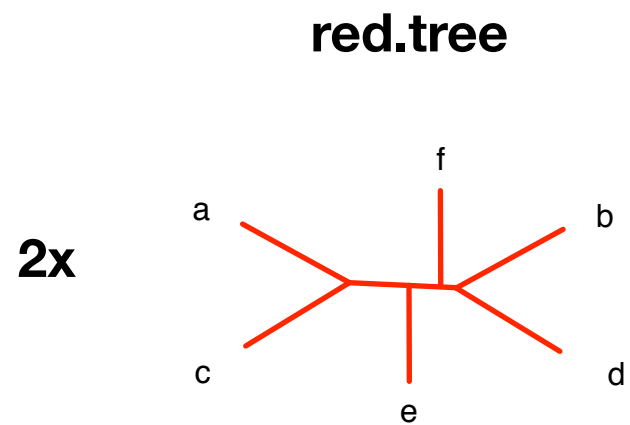
DTL
species tree



abcdef/

Combine the three trees into the same file to have 2x red.tree, 3x green.tree, 5x blue.tree

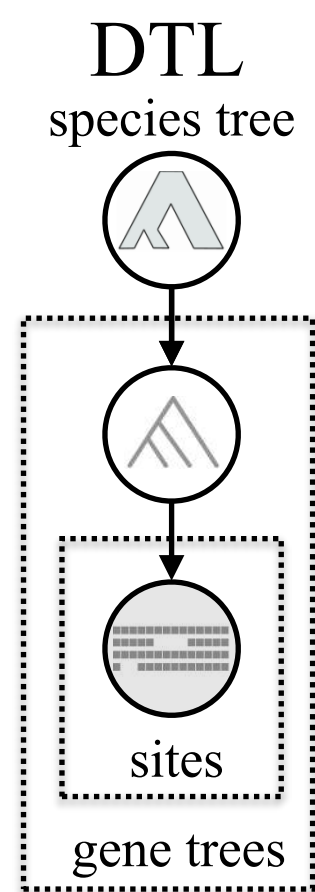
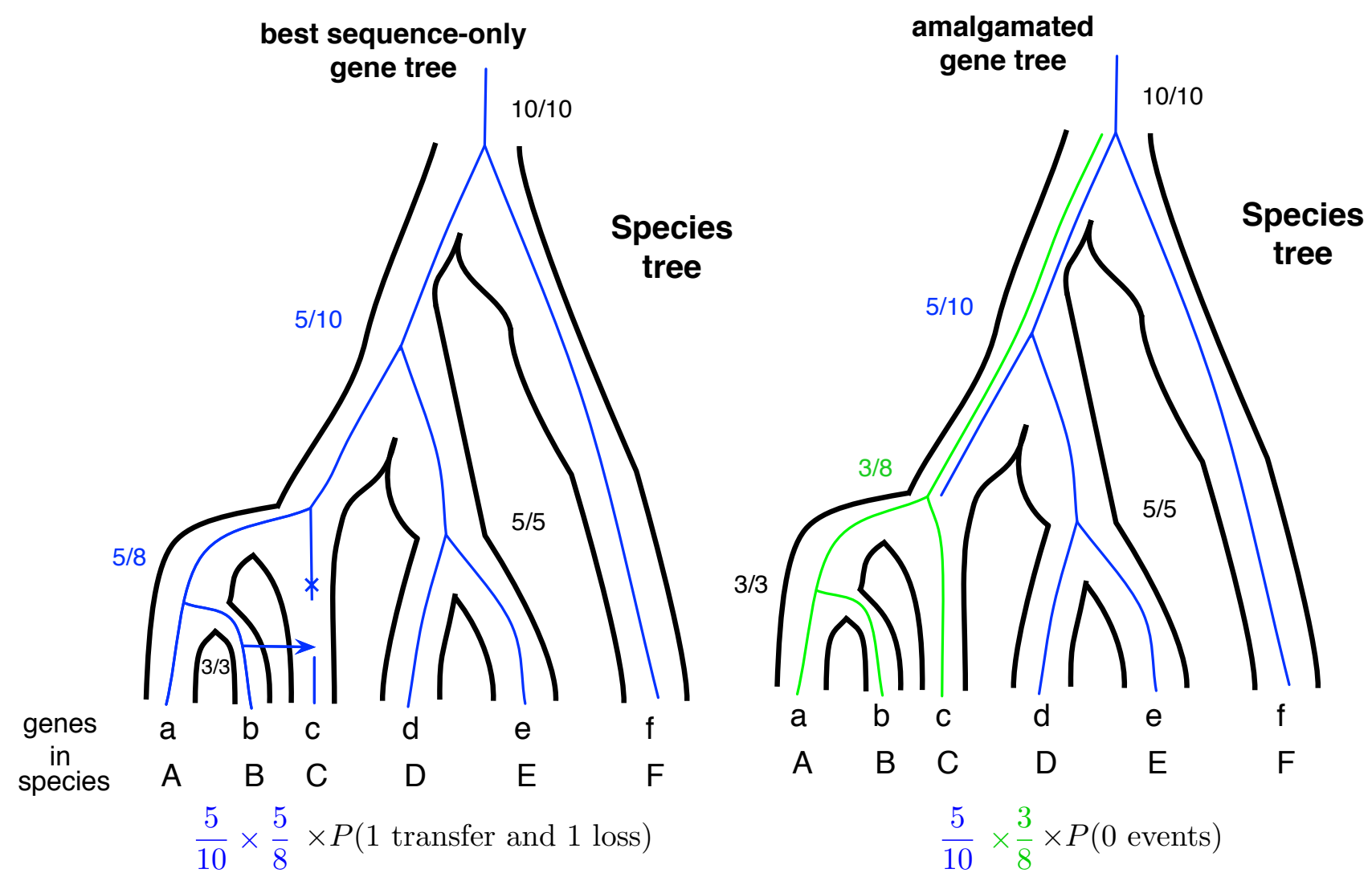
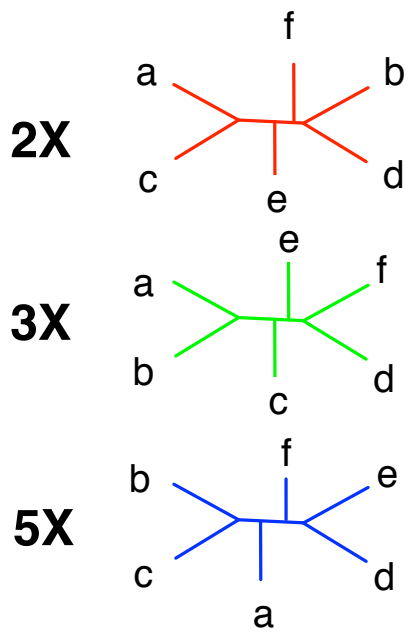
Then use ALEml (and/or ALEml_undated) on each gene tree to estimate ML rates
and sample reconciliations with the specie tree



abcdef/

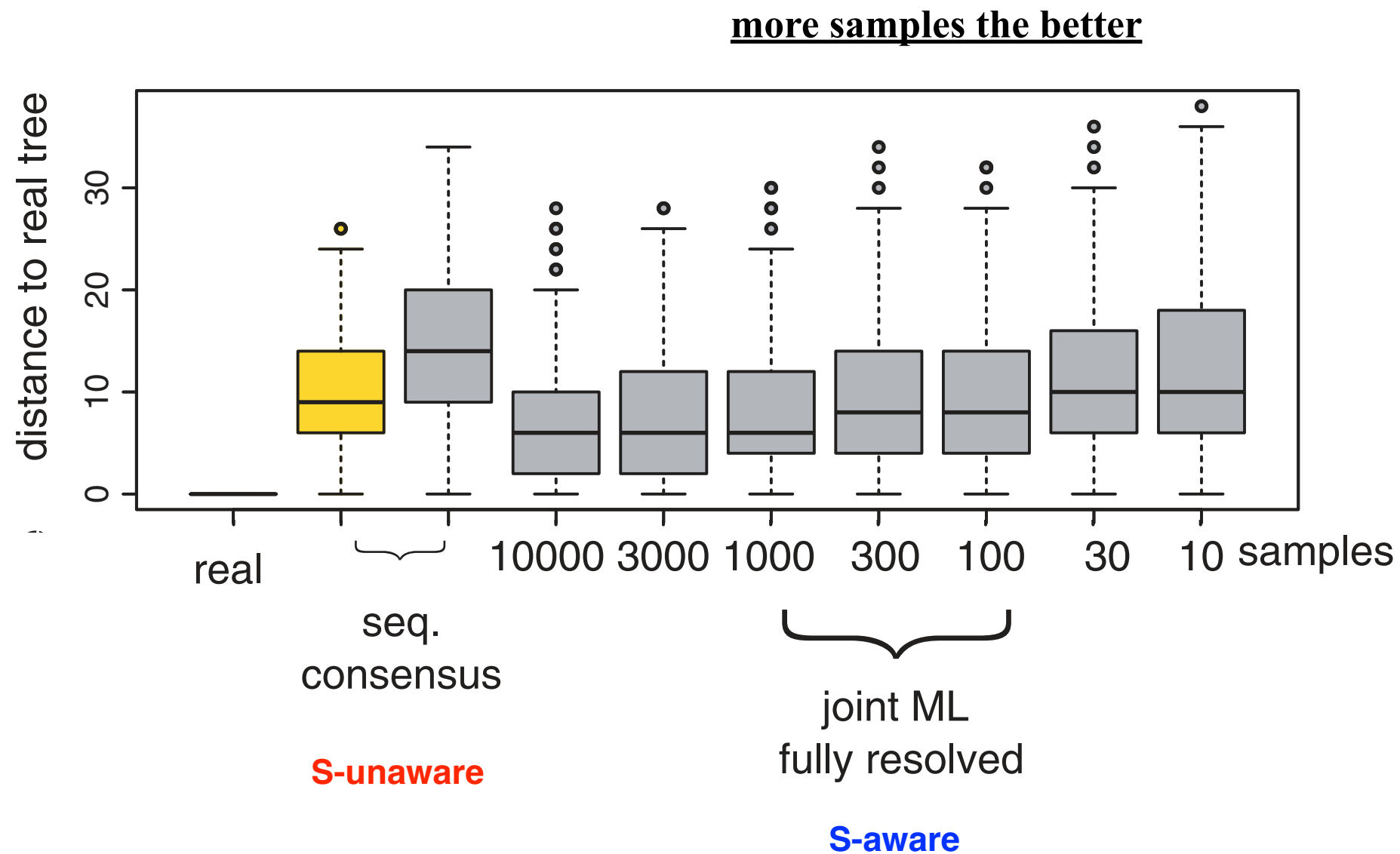
ALEml S.tree help/rgb.trees.ale

sample of gene trees
based on sequence



Efficiently exploring the space of reconciled gene trees

Based on a sample of trees conditional clade probabilities can be used to estimate posterior probability of any gene tree that can be amalgamated. This is usually a very large number of trees (e.g. for 10^4 samples 10^{12} trees, but up to 10^{40}).

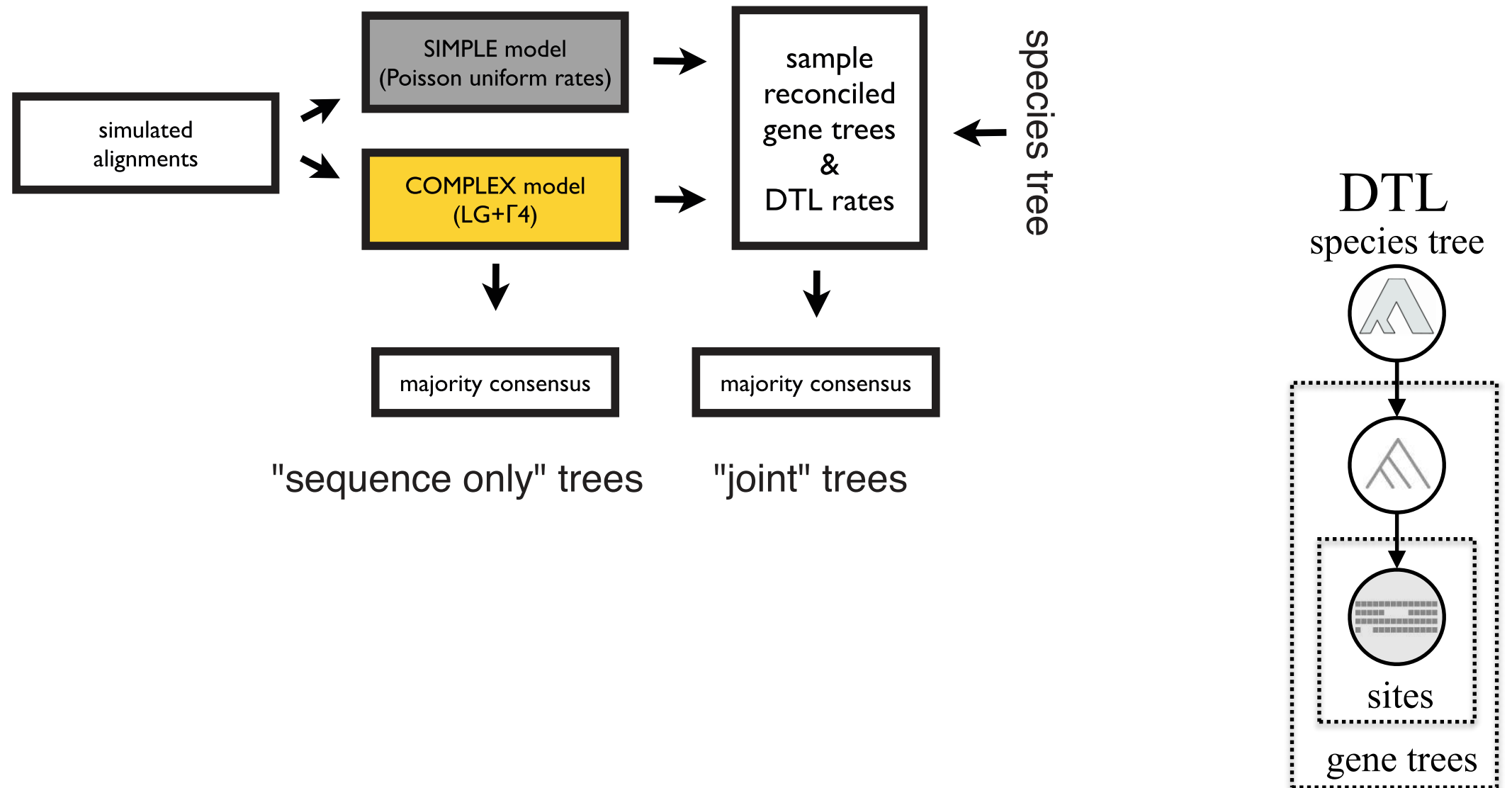


real_data/simulated/sc_univ_fams

Real data!

OK .. first realistic data

```
$ cd ~/workshop_materials/lab_data/real_data/simulated
$ cd sc_univ_fams
```



real_data/simulated/sc_univ_fams

Real data!

OK .. first realistic data

Sample trees using bootstrap and take a look:

```
iqtree2 -m LG -bb 10000 -s HBG486560_sim.fasta -wbtl  
less HBG486560_sim.fasta.ufboot
```

Run ALEml_undated on the ML tree

```
ALEobserve HBG486560_sim.fasta.treefile  
ALEml_undated ../S.tree HBG486560_sim.fasta.treefile.ale
```

Run ALEml_undated on the **true** tree

```
ALEobserve HBG486560_true.tree  
ALEml_undated ../S.tree HBG486560_true.tree.ale
```

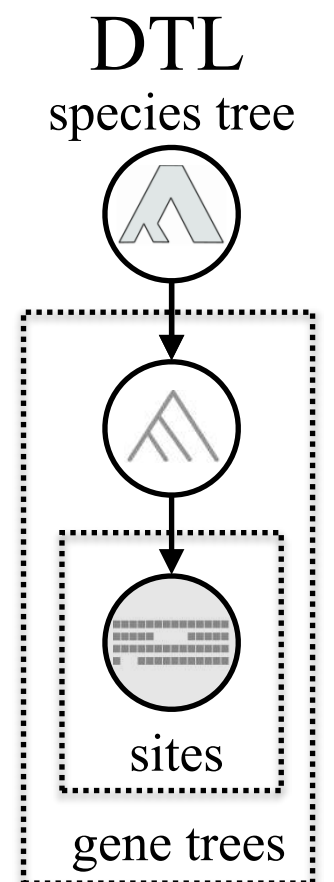
Run ALEml_undated on the **sample of trees**

```
ALEobserve HBG486560_sim.fasta.ufboot  
ALEml_undated ../S.tree HBG486560_sim.fasta.ufboot.ale
```

Do the same for HBG747311_sim.fasta..

What do you see?

(Hint: compare the “Total” rows giving the avg. number of events)



real_data/sc_univ_fams

Real data!

```
$ cd ~/workshop_materials/lab_data/real_data/  
$ cd sc_univ_fams
```

Run ALEml_undated on the ML tree

```
ALEml_undated ../S.tree HBG486560_real.fasta.treefile.ale
```

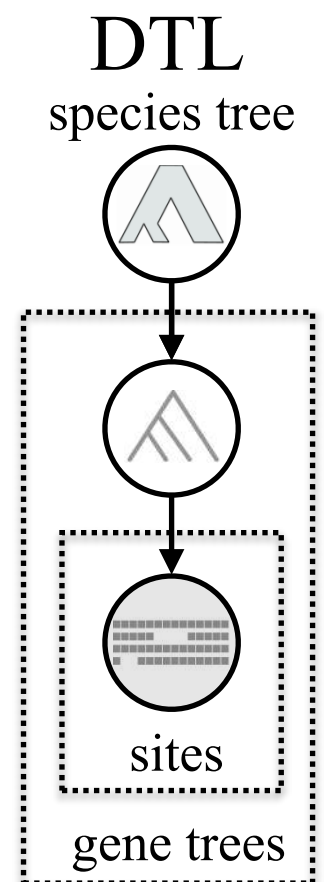
we don't know the true tree, sorry

Run ALEml_undated on the **sample of trees** from phylobayes

```
ALEml_undated ../S.tree HBG486560_real.ale
```

What do you see?

(Hint: compare the “Total” rows giving the avg. number of events)



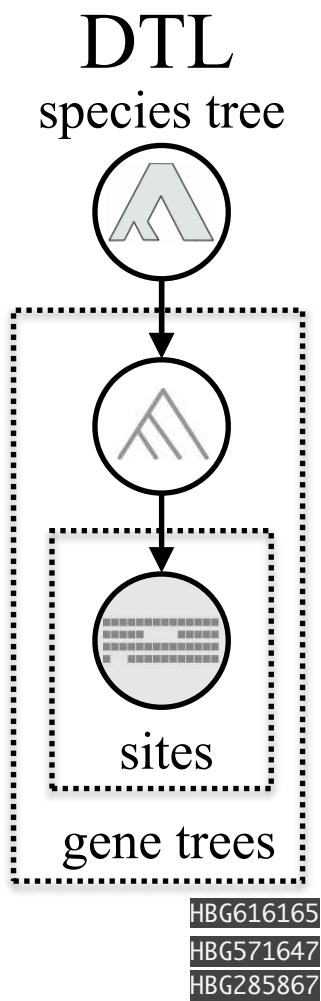
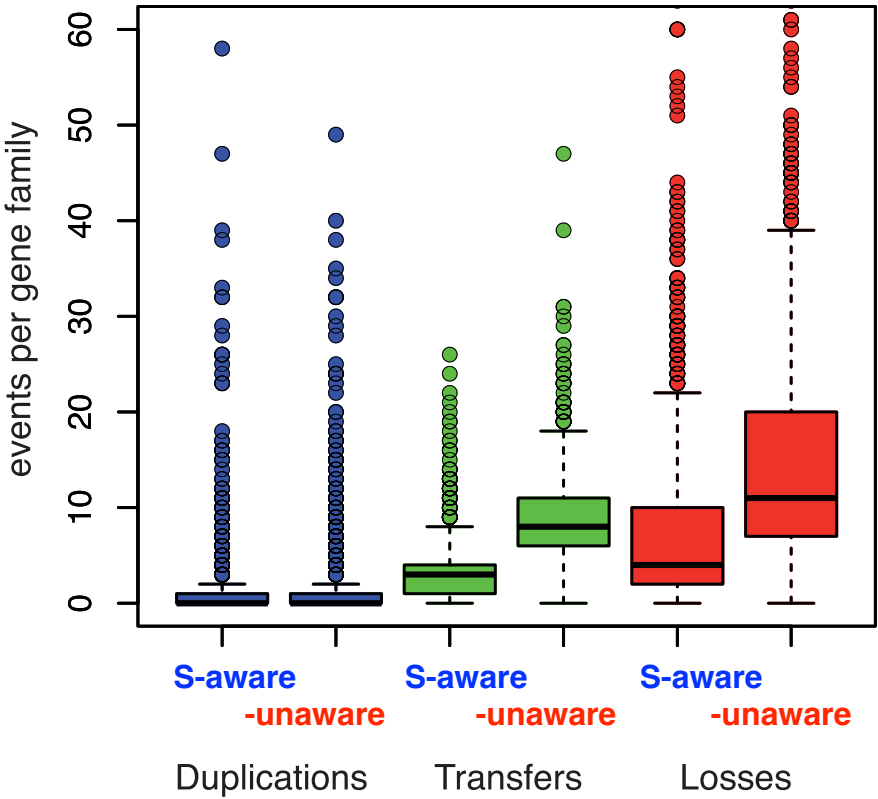
Real data!

ALEml_undated ../S.tree HBG486560_real.fasta.treefile.ale

# of	Duplications	Transfers	Losses	Speciations
Total	0	6.25	3.96	32.71

ALEml_undated ../S.tree HBG486560_real.ale

# of	Duplications	Transfers	Losses	Speciations
Total	0	1.05	0.99	34.94



real_data/sc_univ_fams

Real data!

ALEml_undated ../S.tree HBG486560_real.fasta.treefile.ale

# of	Duplications	Transfers	Losses	Speciations
Total	0	6.25	3.96	32.71

ALEml_undated ../S.tree HBG486560_real.ale

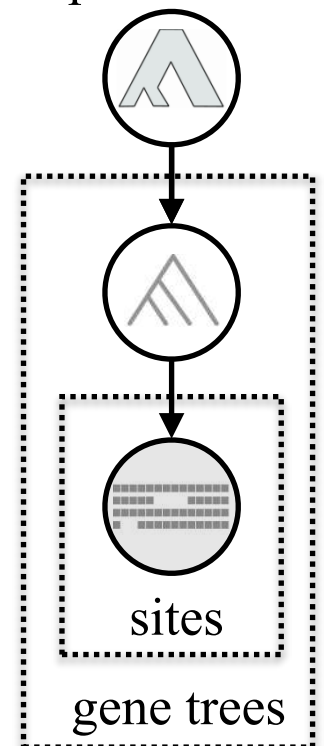
# of	Duplications	Transfers	Losses	Speciations
Total	0	1.05	0.99	34.94

Can you find a true orthologous family?
(i.e. a family with 0 Duplication , Transfer and Loss events)

Hint: if you are less familiar with the command line you can ask ChatGPT or Bard how to run the same command on multiple files?

```
# Loop through each .ale file in the current directory
for ale_file in *.ale; do
    echo "Processing $ale_file..."
    ALEml_undated ../S.tree "$ale_file"
done
```

DTL
species tree



HBG616165
HBG571647
HBG285867

real_data/general_fams

Real data!

```
$ cd ~/workshop_materials/lab_data/real_data/simulated
$ cd general_fams
```

Find the rooted species tree with the highest likelihood using DTL by summing the log-likelihood across all .ale files (cf. bash command on previous page) for each candidate species tree S.tree, S_alt1.tree, S_alt2.tree (or your own rerooted version of S.tree made with seaview or FigTree).

```
ALEml_undated ../S.tree $ale_file
```

```
$ wget http://ssolo.web.elte.hu/S_alt1.tree
```

```
ALEml_undated S_alt1.tree $ale_file
```

```
$ wget http://ssolo.web.elte.hu/S_alt2.tree
```

```
ALEml_undated S_alt2.tree $ale_file
```

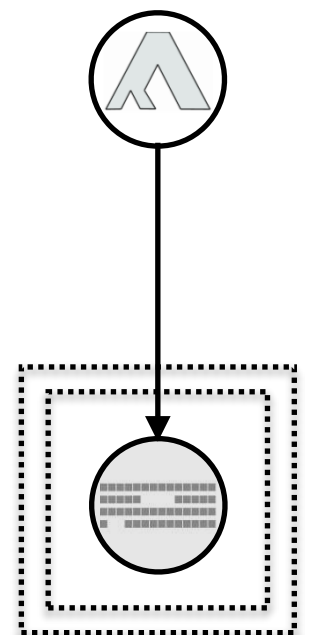
Bonus/Home Work:

try our new implementation that can do this and a lot more for you
(efficiently using MPI):

<https://github.com/BenoitMorel/AleRax.git>

DTL

species tree





Join us for a postdoc at the interface of computational & evolutionary biology! Use probabilistic models & machine learning to model coevolution, reconstruct the Tree of Life, understand somatic evolution or pursue your own project..

Model-Based Evolutionary Genomics Unit
モデルベース進化ゲノミクスユニット
<https://www.oist.jp/research/research-units/modevolgenom>

Okinawa Institute of Science and Technology

gergely.szollosi@oist.jp

Bonus/Home Work:
try our new implementation that can do this and a lot more for you
(efficiently using MPI):

<https://github.com/BenoitMorel/AleRax.git>