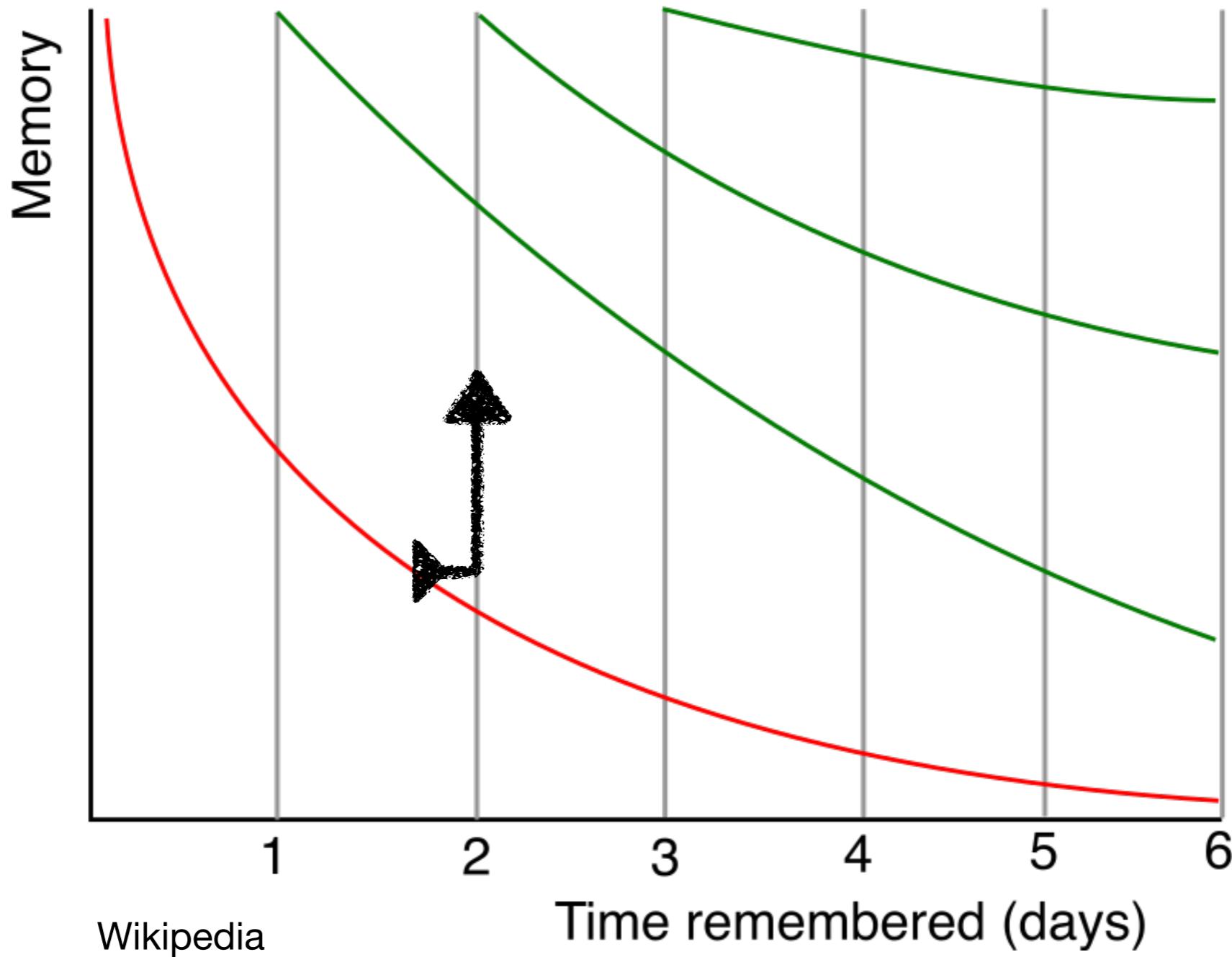


Last, but not least

The Forgetting Curve



Reconciliation Workshop

Oliver Schick and Lenard Szanho

Model-Based Evolutionary Genomics Unit

McGlynn Lab visiting OIST

~~May 14, May 15, May 16, 2025~~



Preparation

- Notebook with Linux or MacOS

- g++, cmake, git, FigTree, openmpi

- Workshop datasets (125 MB, approx 4 minutes download):

```
$ cd ~  
$ git clone https://github.com/oist/reconciliation-workshop.git  
$ cd reconciliation-workshop  
$ git pull
```

- AleRax software (version 1.2.1 included in the git repository, 2 mins to compile):

```
$ cd AleRaxRep  
$ ./install.sh  
$ cd ..
```

- ALE software (9 mins to compile)

```
$ brew install openmpi // sudo apt install openmpi-bin libopenmpi-dev  
$ git clone https://github.com/ssolo/ALE.git  
$ cd ALE; mkdir build; cd build  
$ cmake ..  
$ make -j16
```

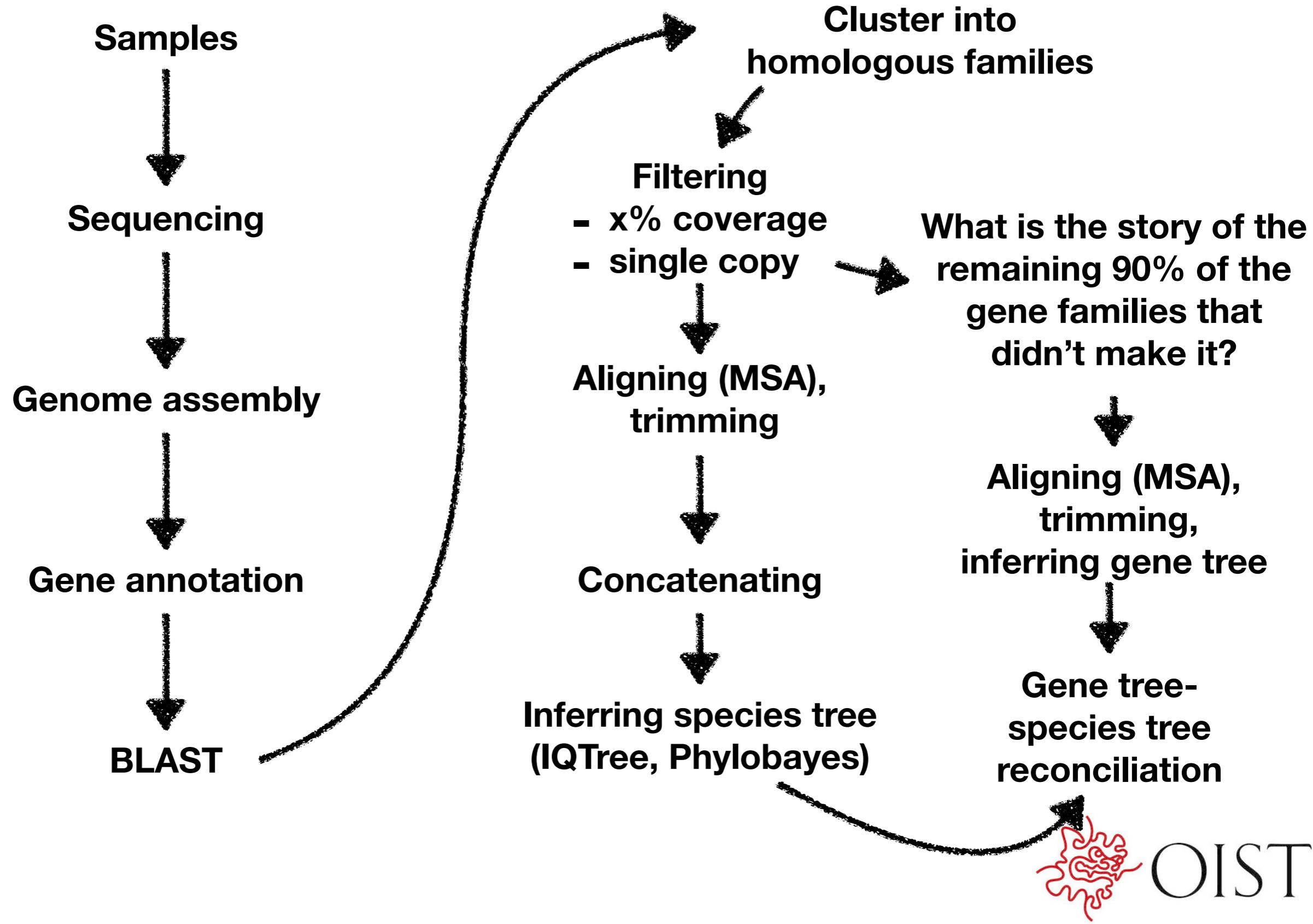
- FigTree:

for MacOS download the pre-release: https://github.com/rambaut/figtree/releases/download/v1.4.5pre/FigTree.v1.4.5_pre.dmg

for Linux:

```
$ sudo apt install figtree
```

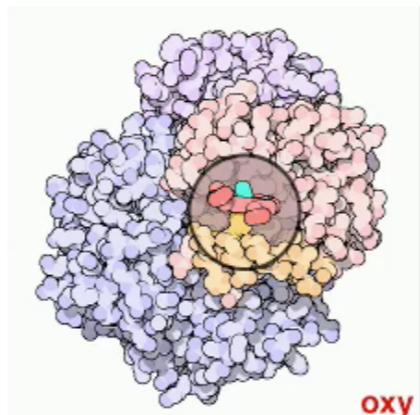
Example of usual phylogenetic pipeline



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.

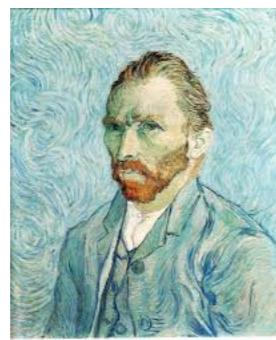
Human hemoglobin is composed of



$2\alpha + 2\beta$ chains.

molecular machine

Human



$2\alpha +$

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

adult



fetus

Cow



adult



fetus

Horse



$+2\{\beta\delta\}$
adult and fetus

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

$+2\gamma$

Hemoglobin/

Hemoglobin gene tree

Hemoglobin.tree



Hemoglobin.tree

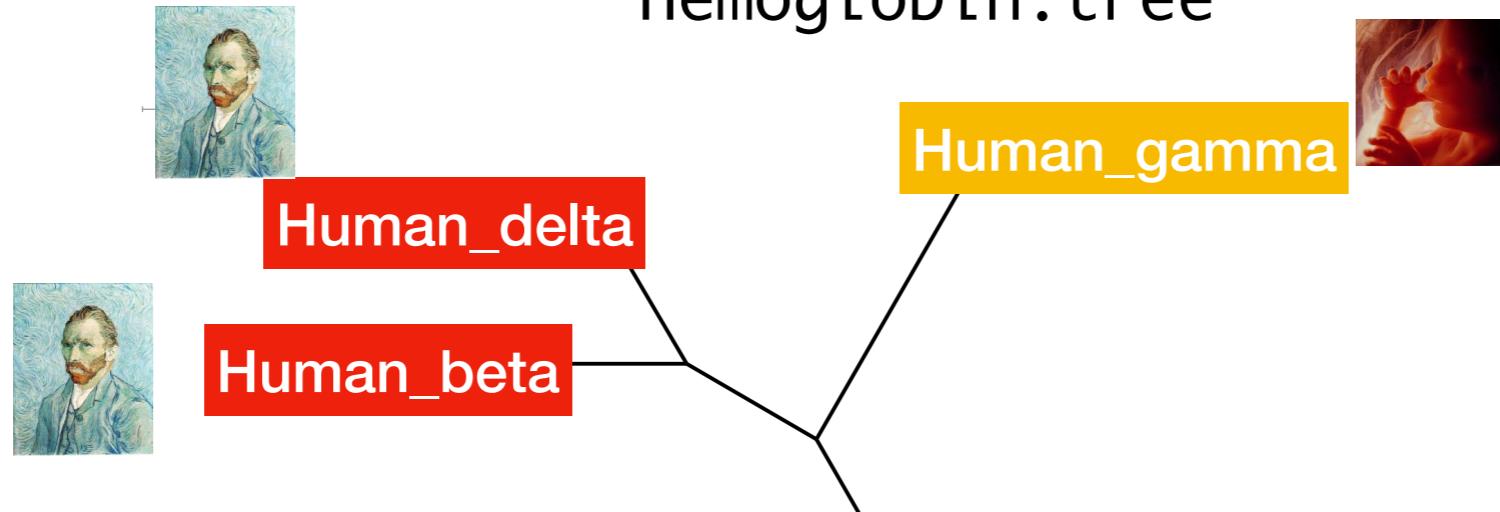
(Human_beta:1,Human_delta:1):1,



Hemoglobin/

Hemoglobin gene tree

Hemoglobin.tree



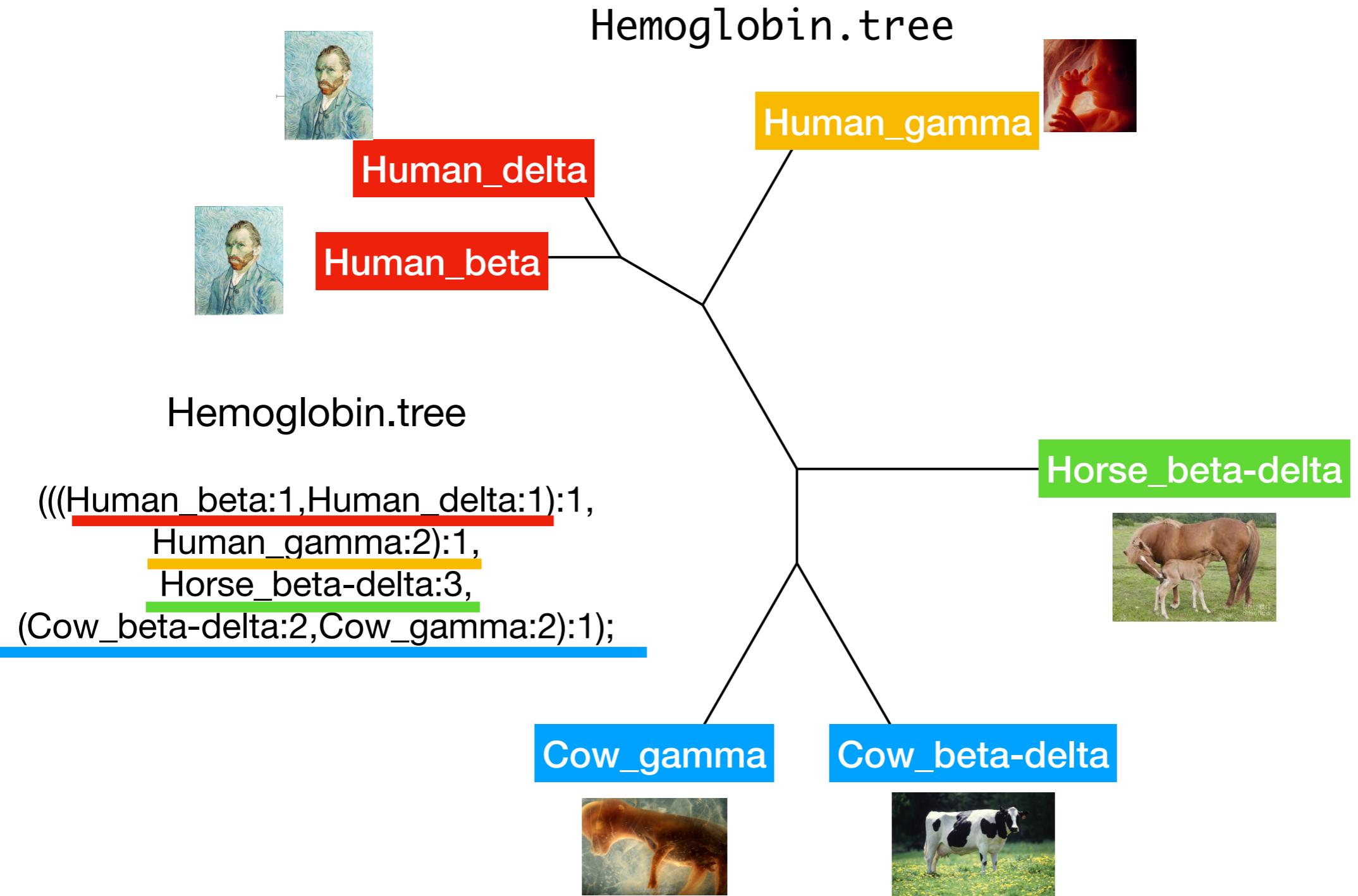
Hemoglobin.tree

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



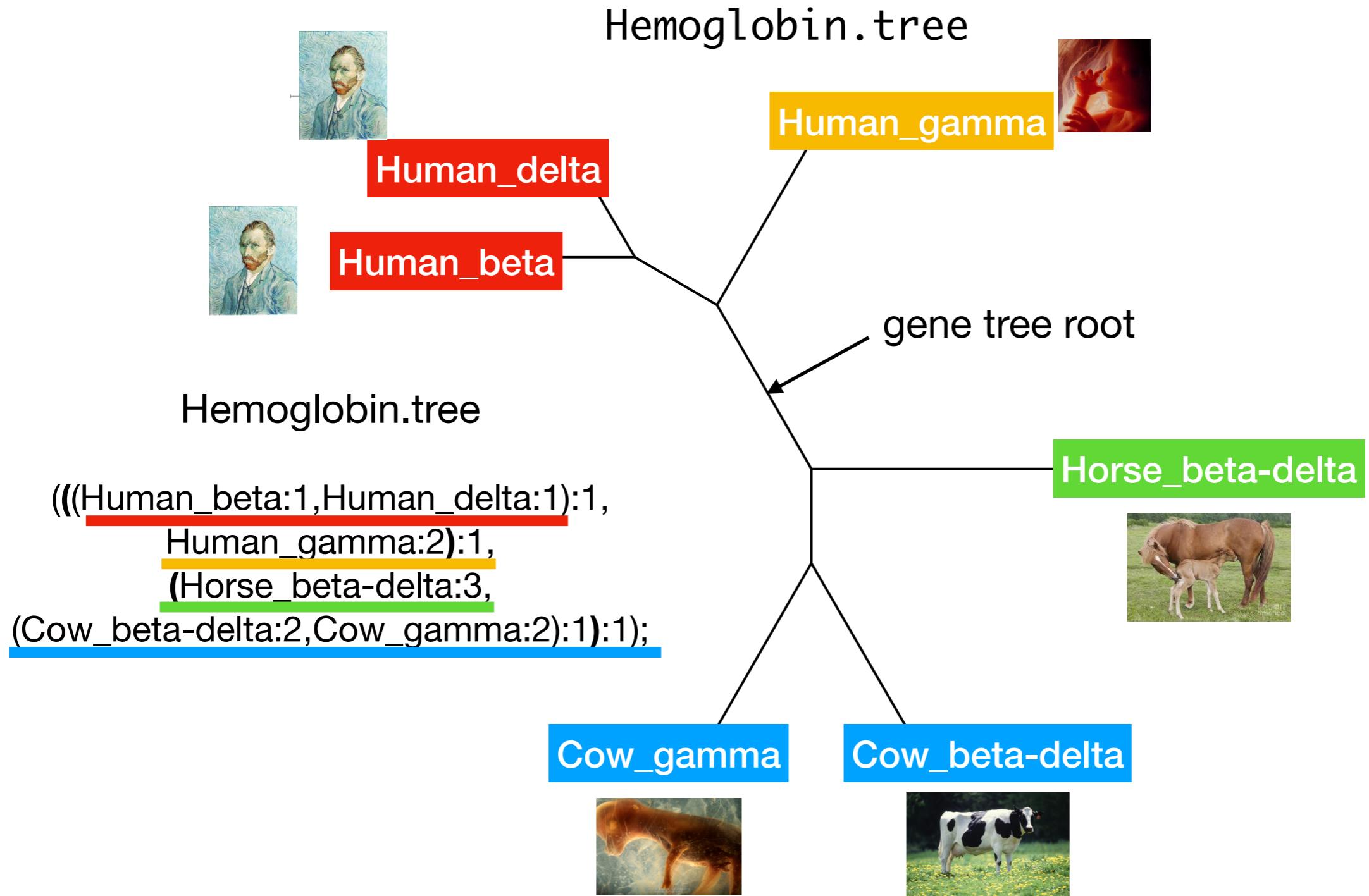
Hemoglobin/

Hemoglobin gene tree



Hemoglobin/

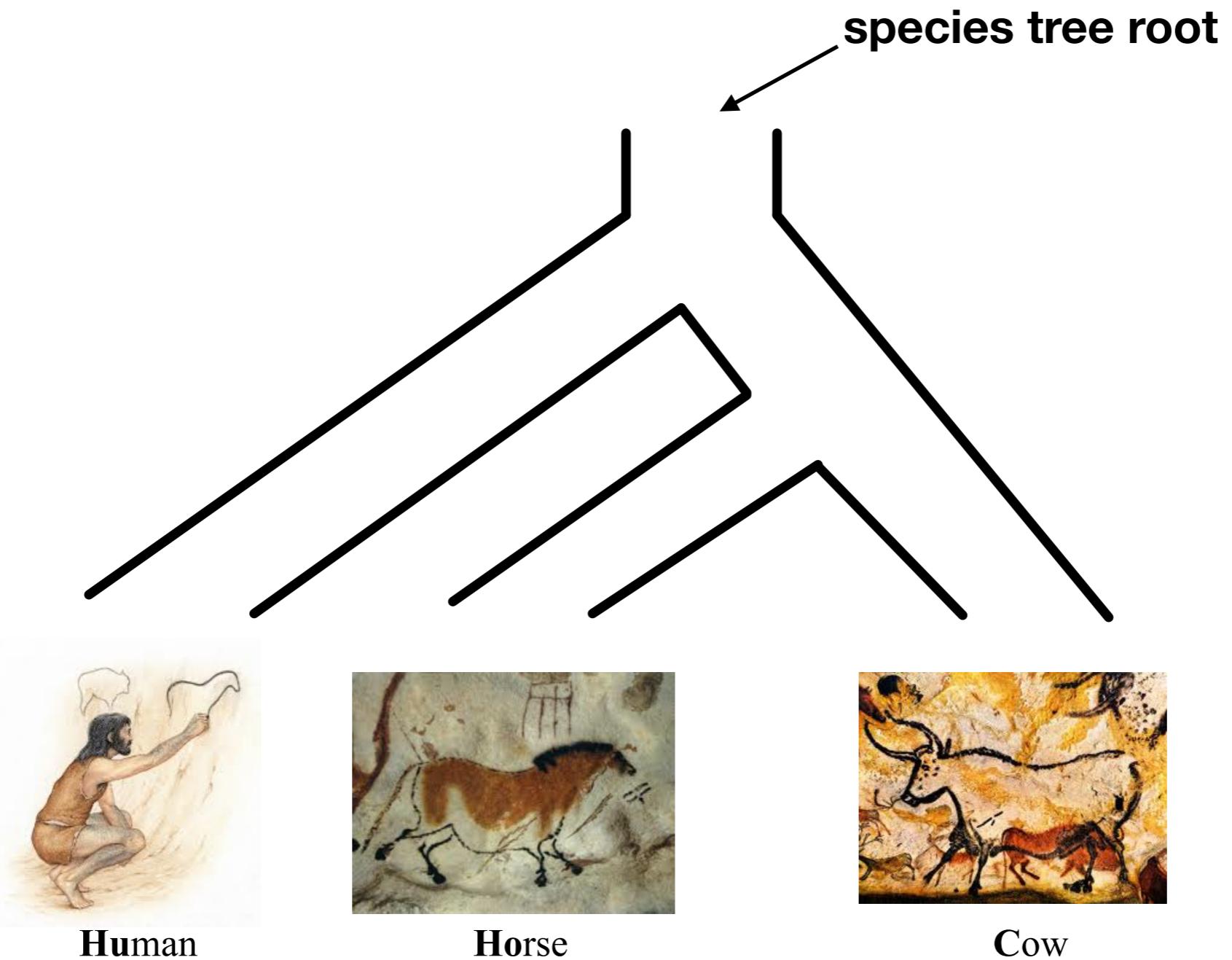
Hemoglobin gene tree



Hemoglobin/

Species tree

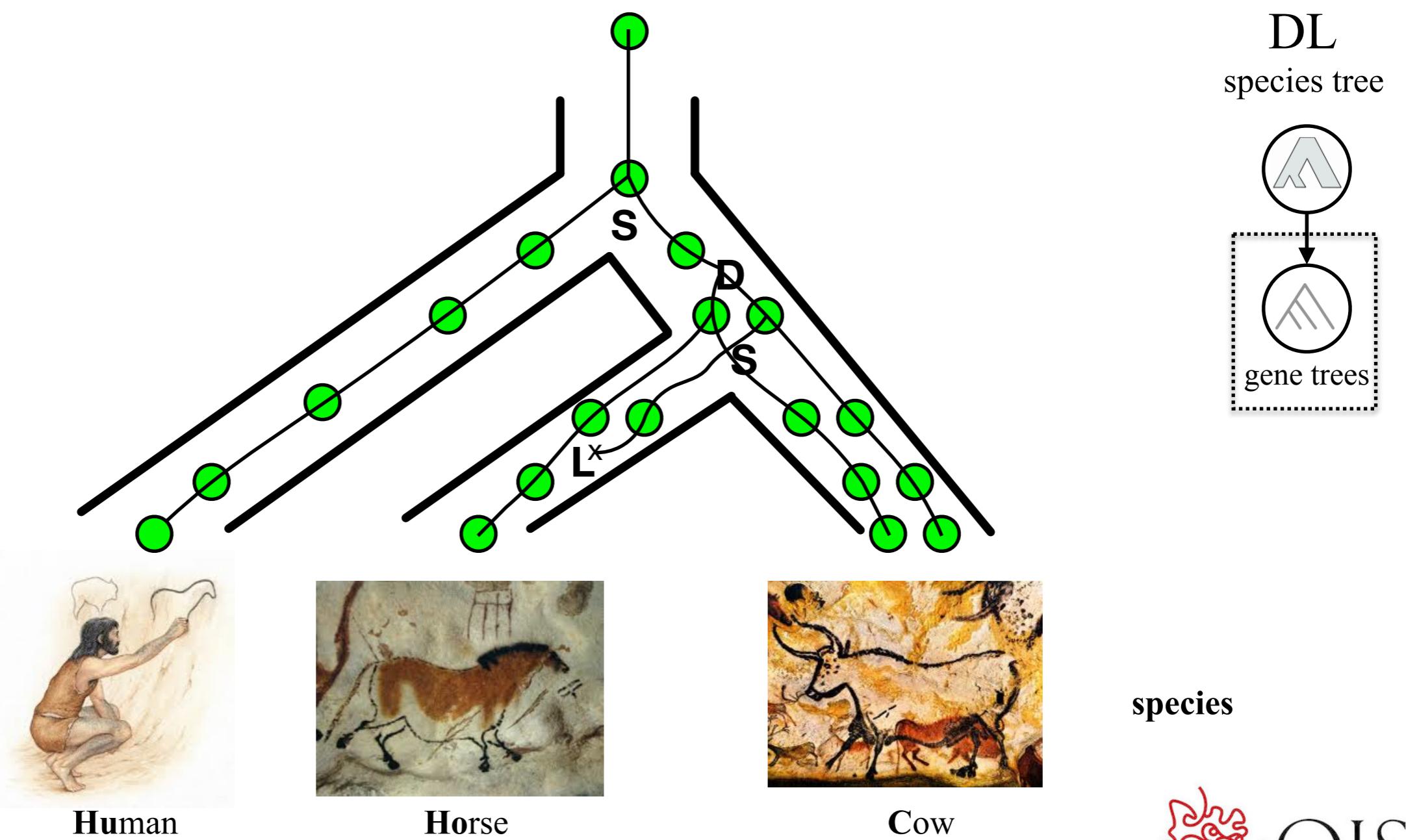
data/HuHoCo.sptree
(Human:2,(Horse:1,Cow:1):1);



species

Hemoglobin/

DL along S



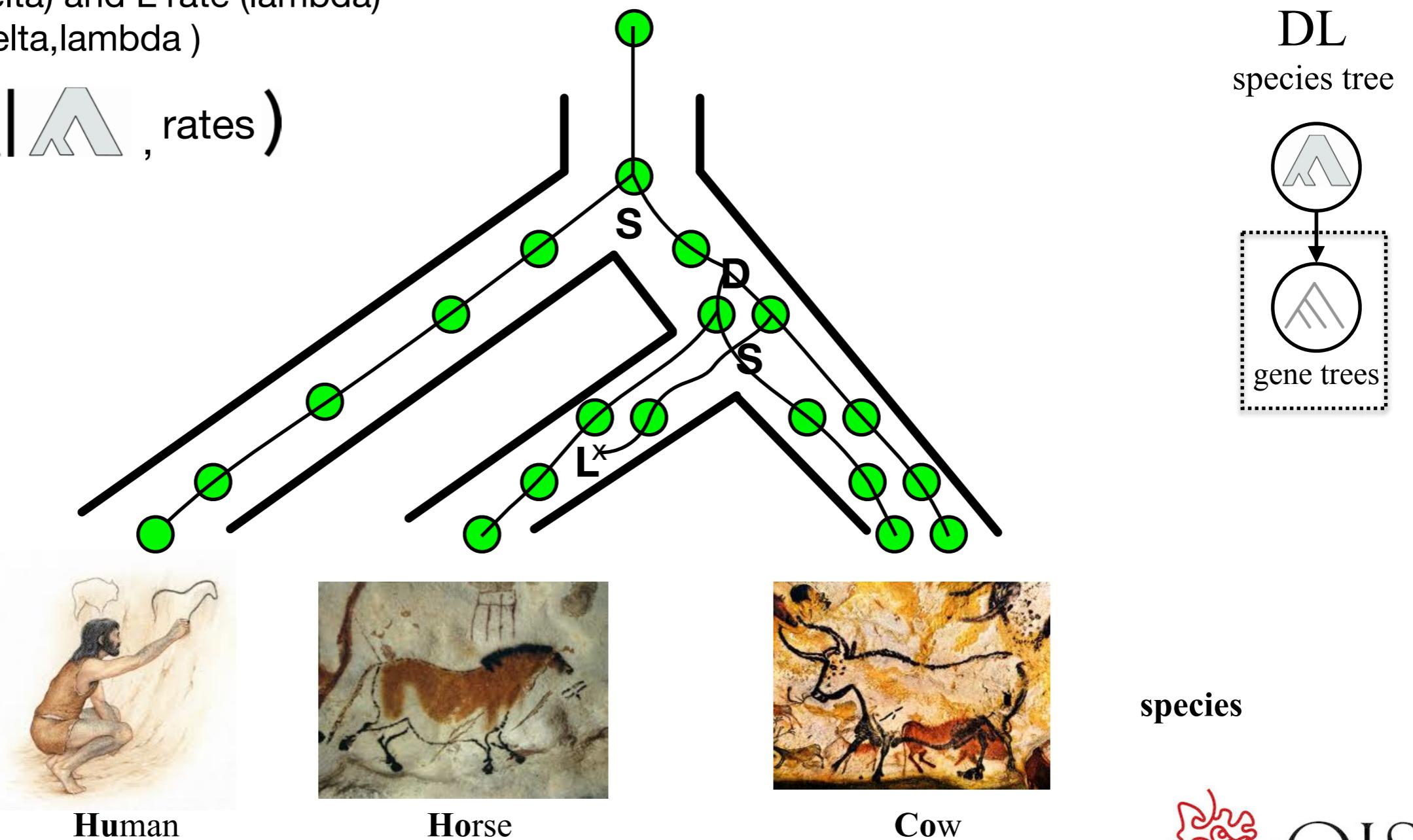
Hemoglobin/

DL along S

Given G and S we want to calculate:

1. D rate (delta) and L rate (lambda)
2. $P(G | S, \delta, \lambda)$

$P(\text{A} \setminus \text{A} | \text{A}, \text{rates})$

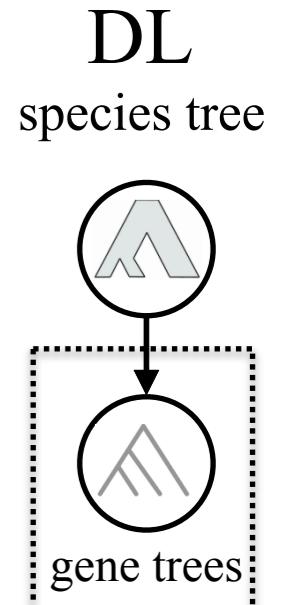


```
Hemoglobin/
$ ./alerax --families Hemoglobin.families
--species-tree data/HuHoCo.sptree --prefix Hemoglobin
--gene-tree-samples 10 --rec-model UndatedDL
--model-parametrization GLOBAL
```

species tree: Hemoglobin/species_trees/starting_species_tree.newick
(Human:2.000000,(Horse:1.000000,Cow:1.000000)Node_Horse_Cow_0:1.000000)Node_Human_Horse_0:1.000000;

ML log-likelihood (DL model): Hemoglobin/per_family_likelihoods.txt
Hemoglobin -8.05492

ML rates (DL model): Hemoglobin/model_parameters/model_parameters.txt
Human 0.375001 1e-10



10 random reconciled gene trees **sampled according to joint likelihood**
Hemoglobin/reconciliations/all/Hemoglobin.rec.uml

((Human_gamma:2,(Human_beta:1,Human_delta:1).D@Human:1).D@Human:2,(Horse_beta-delta:3,(Cow_gamma:2,Cow_beta-delta:2).D@Cow:1).Node_Horse_Cow_0:2).Node_Human_Horse_0;
((Human_gamma:2,(Human_beta:1,Human_delta:1).D@Human:1).D@Human:2,(Horse_beta-delta:3,(Cow_gamma:2,Cow_beta-delta:2).D@Cow:1).Node_Horse_Cow_0:2).Node_Human_Horse_0;

mean number of events per branch in sampled
reconciliations: Hemoglobin/reconciliations/
perspecies_eventcount.txt

species_label,	speciations, duplications, losses, transfers, presence, origination, copies, singletons
Cow,	2, 1, 0, 0, 1, 0, 2, 2
Horse,	1, 0, 0, 0, 1, 0, 1, 1
Human,	3, 2, 0, 0, 1, 0, 3, 3
Node_Horse_Cow_0,	1, 0, 0, 0, 1, 0, 1, 1
Node_Human_Horse_0,	1, 0, 0, 0, 1, 1, 1, 1

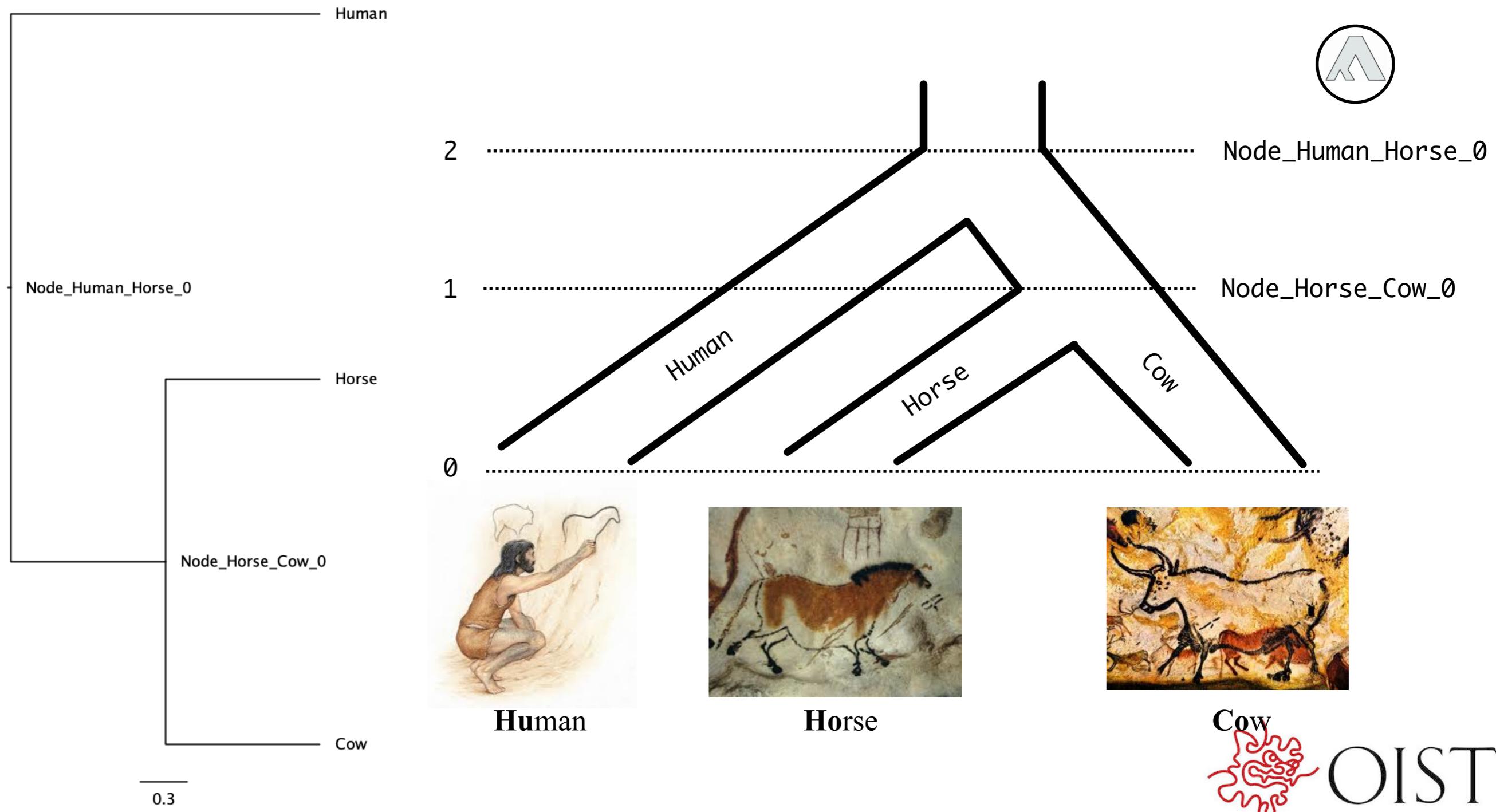
$$\sum P(\text{=====| } \Delta) P(\Delta | \Delta)$$

sum over reconciled
gene trees

```

Hemoglobin/
$ ./alerax --families Hemoglobin.families
--species-tree data/HuHoCo.sptree --prefix Hemoglobin
--gene-tree-samples 10 --rec-model UndatedDL
--model-parametrization GLOBAL
Species tree: (Human:2.000000,(Horse:1.000000,Cow:1.000000)Node_Horse_Cow_0:1.000000)Node_Human_Horse_0:1.000000;

```

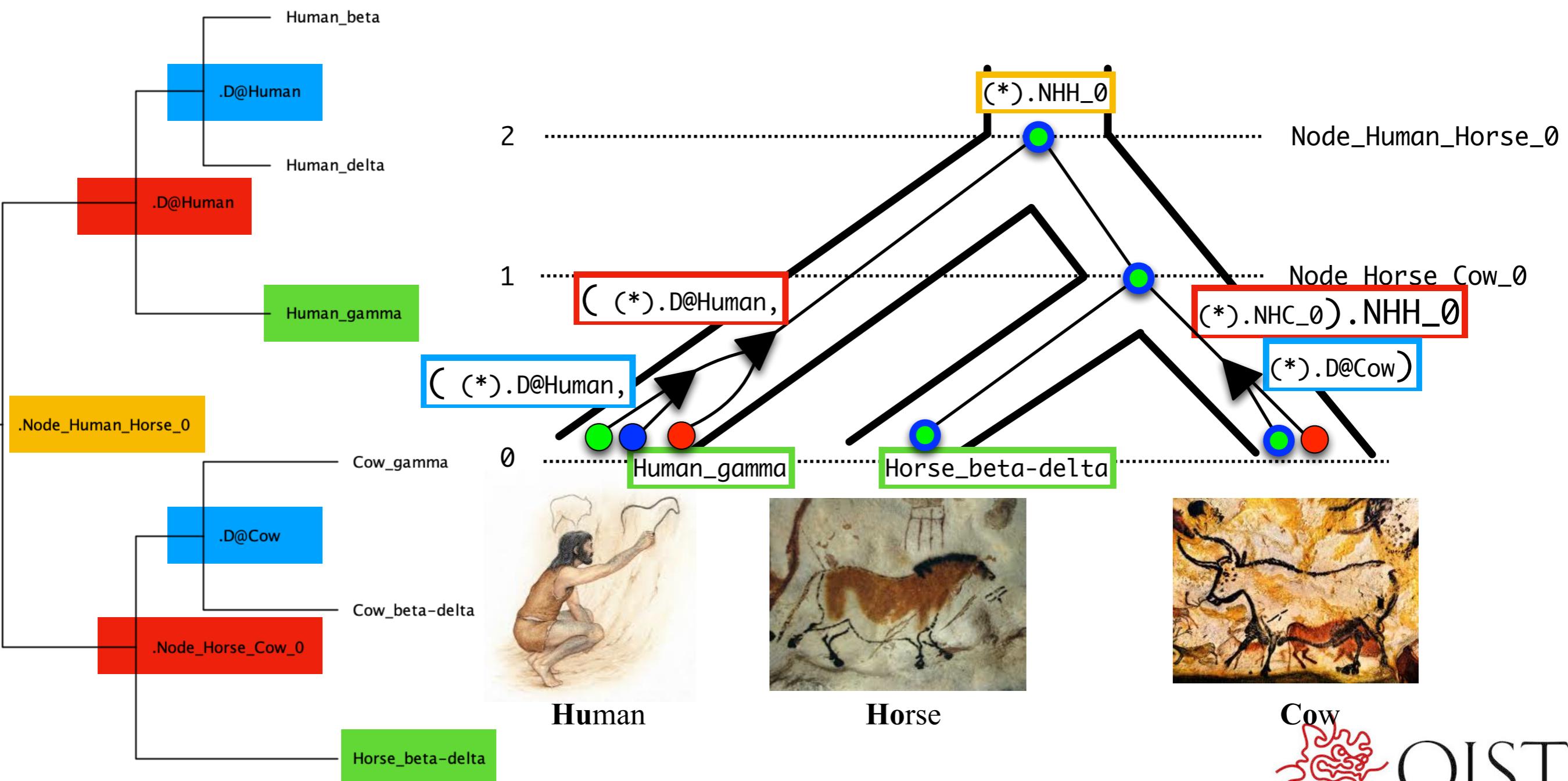


Hemoglobin/

((Human_gamma:2,(Human_beta:1,Human_delta:1).D@Human:1).D@Human:2,(Horse_beta-delta:3,(Cow_gamma:2,Cow_beta-delta:2).D@Cow:1).Node_Horse_Cow_0:2).Node_Human_Horse_0;

((Human_gamma:2,(Human_beta:1,Human_delta:1).D@Human:1).D@Human:2 ,

,(Horse_beta-delta:3 ,(Cow_gamma:2,Cow_beta-delta:2).D@Cow:1).Node_Horse_Cow_0:2).Node_Human_Horse_0:0;

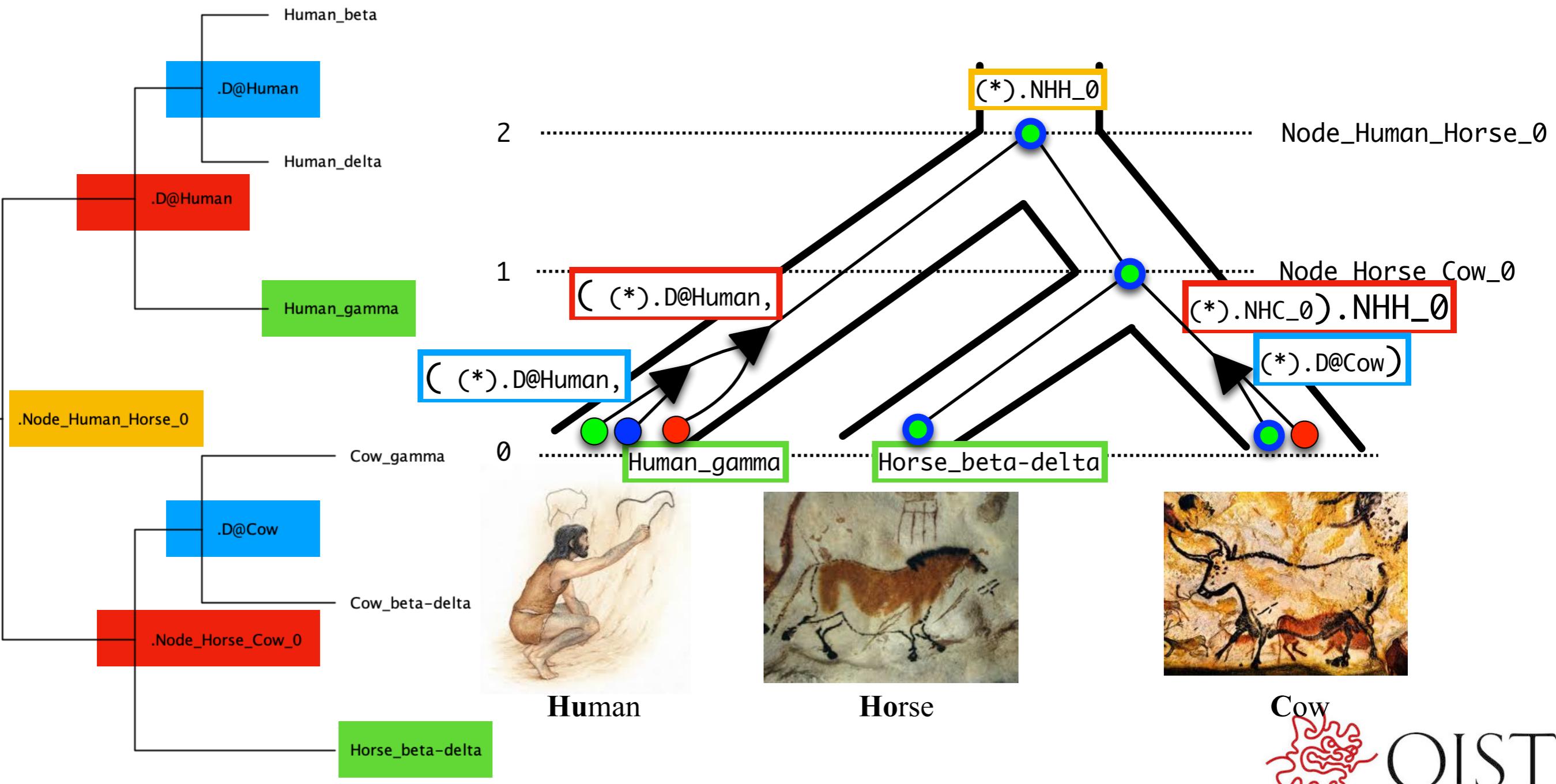


Hemoglobin/

((Human_gamma:2,(Human_beta:1,Human_delta:1).D@Human:1).D@Human:2,(Horse_beta-delta:3,(Cow_gamma:2,Cow_beta-delta:2).D@Cow:1).Node_Horse_Cow_0:2).Node_Human_Horse_0;

mean number of events per branch in sampled reconciliations:

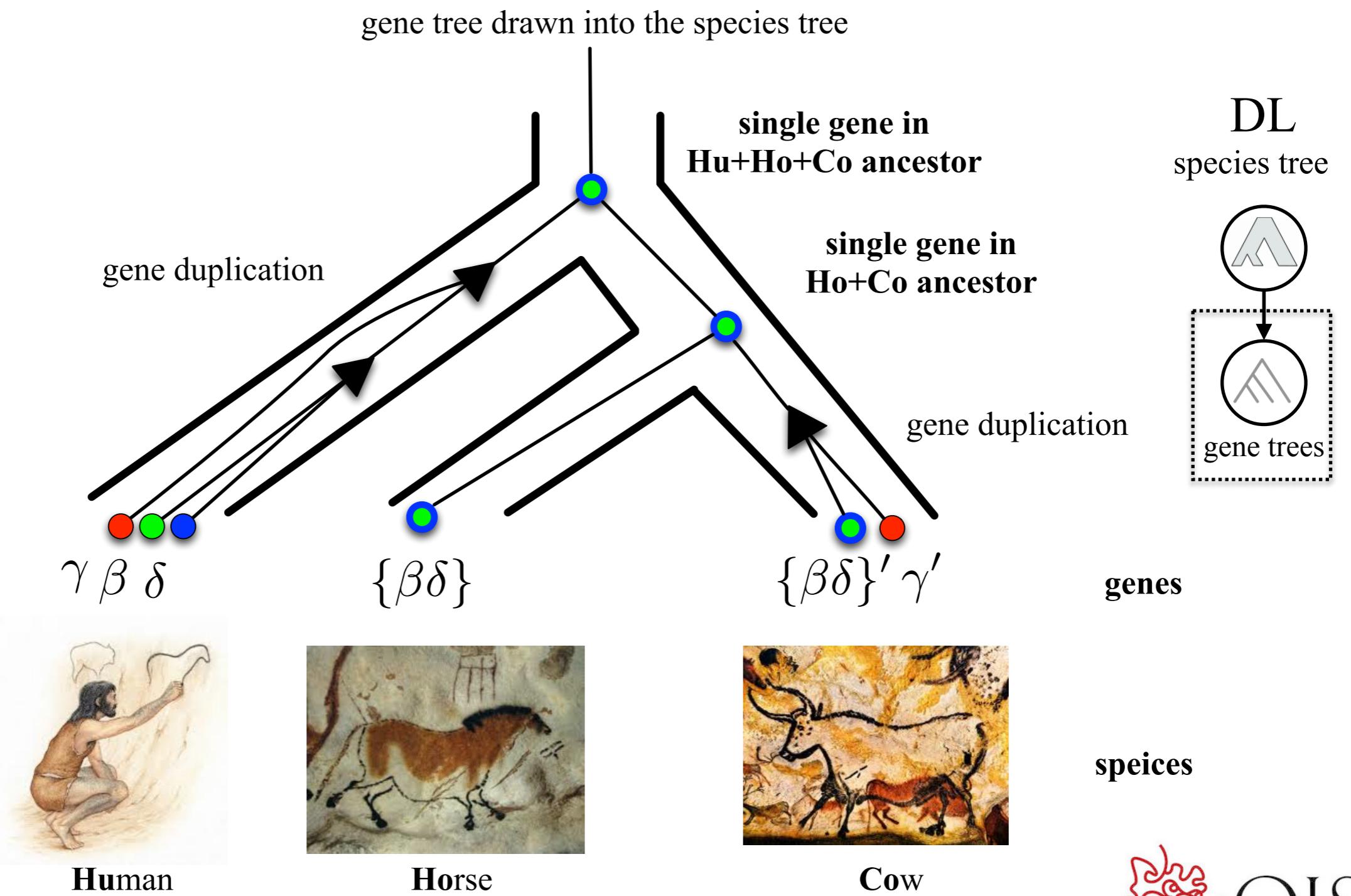
species_label,	speciations,	duplications,	losses,	transfers,	presence,	origination,	copies,	singletons
Cow,	2,	1,	0,	0,	1,	0,	2,	2
Horse,	1,	0,	0,	0,	1,	0,	1,	1
Human,	3,	2,	0,	0,	1,	0,	3,	3
Node_Horse_Cow_0,	1,	0,	0,	0,	1,	0,	1,	1
Node_Human_Horse_0,	1,	0,	0,	0,	1,	1,	1,	1



Hemoglobin/

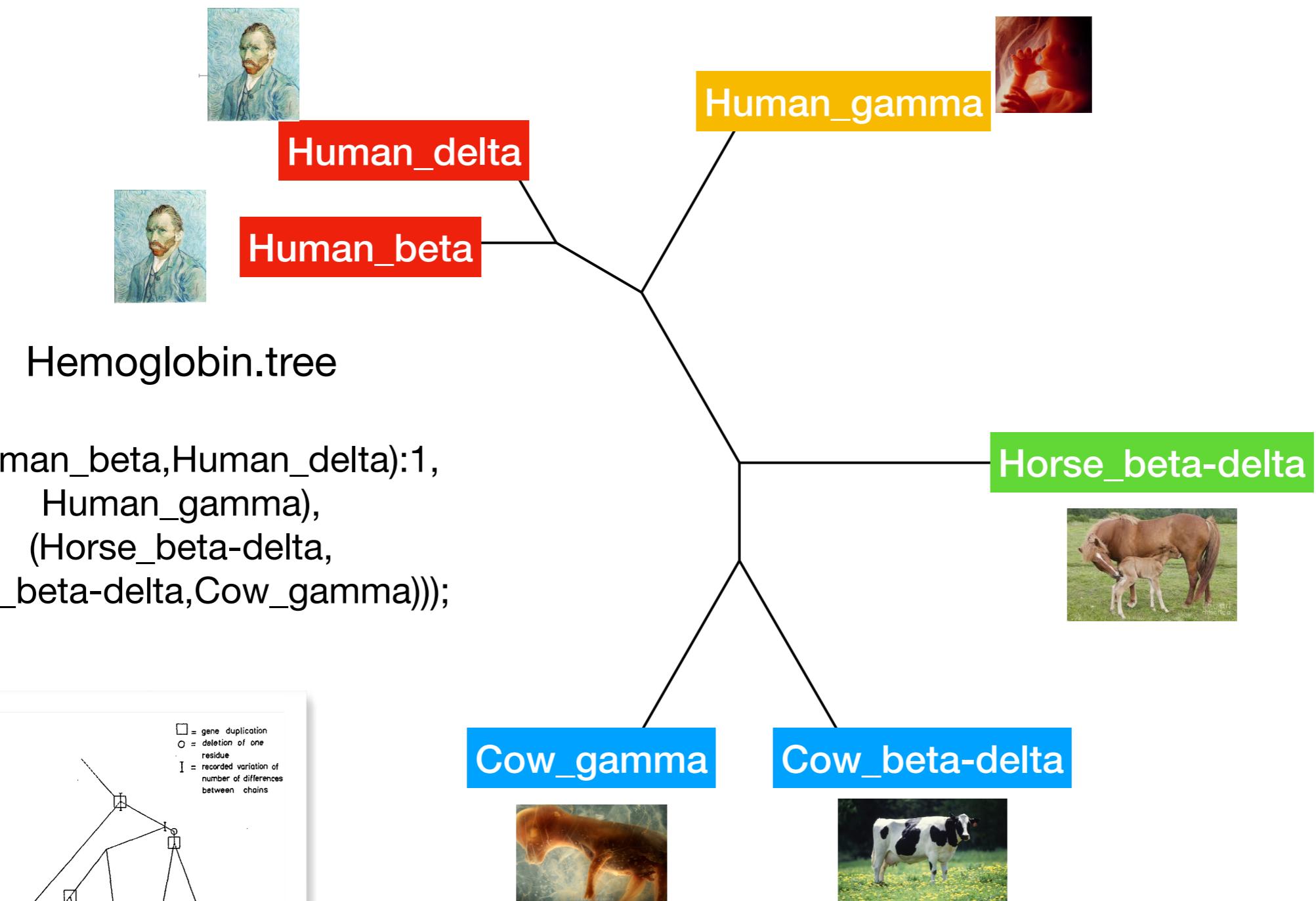
The stories of gene families can be complicated

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.



Hemoglobin/

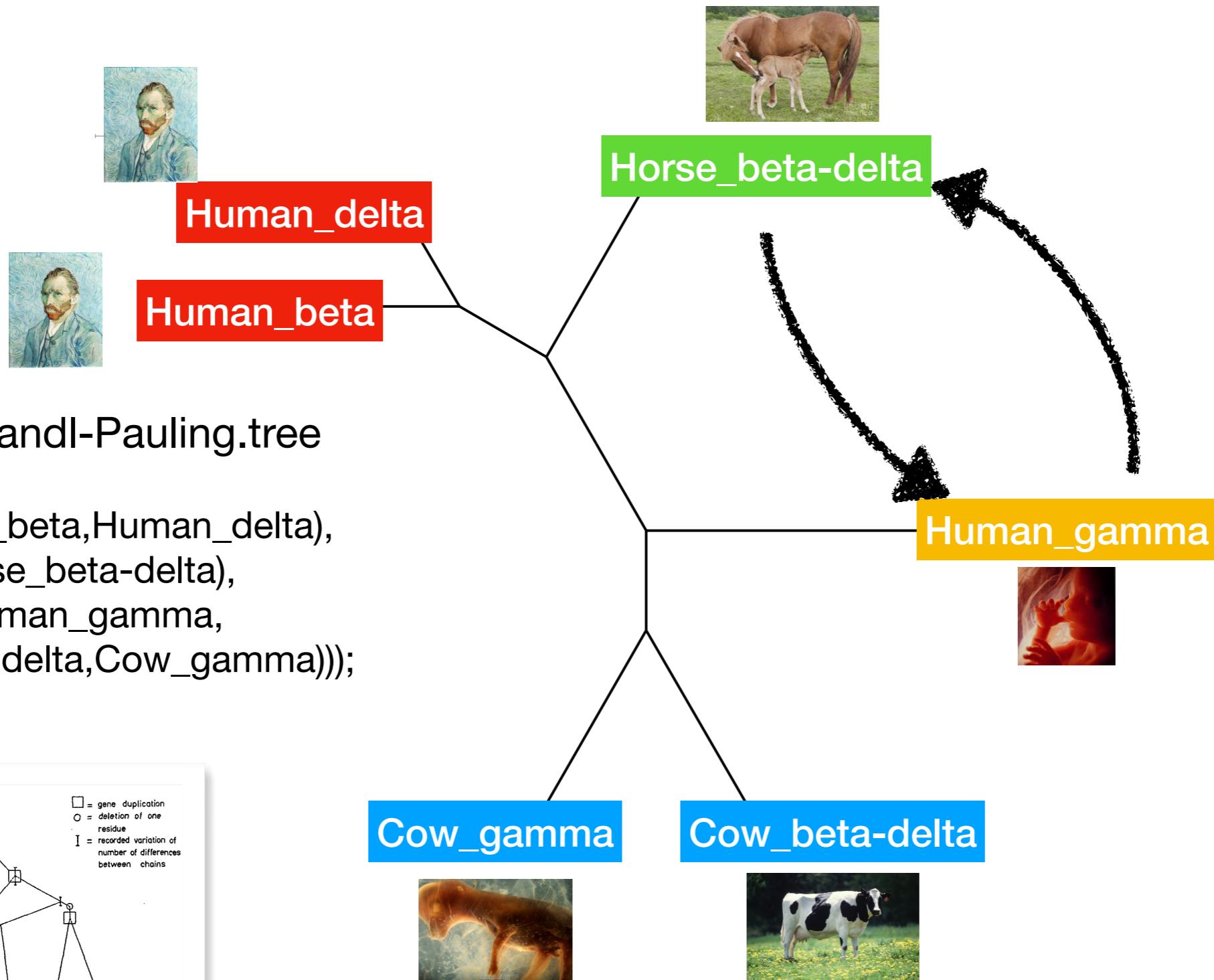
Hemoglobin gene tree



Zuckerkandl & Pauling 1965

Hemoglobin/

The first ever gene tree

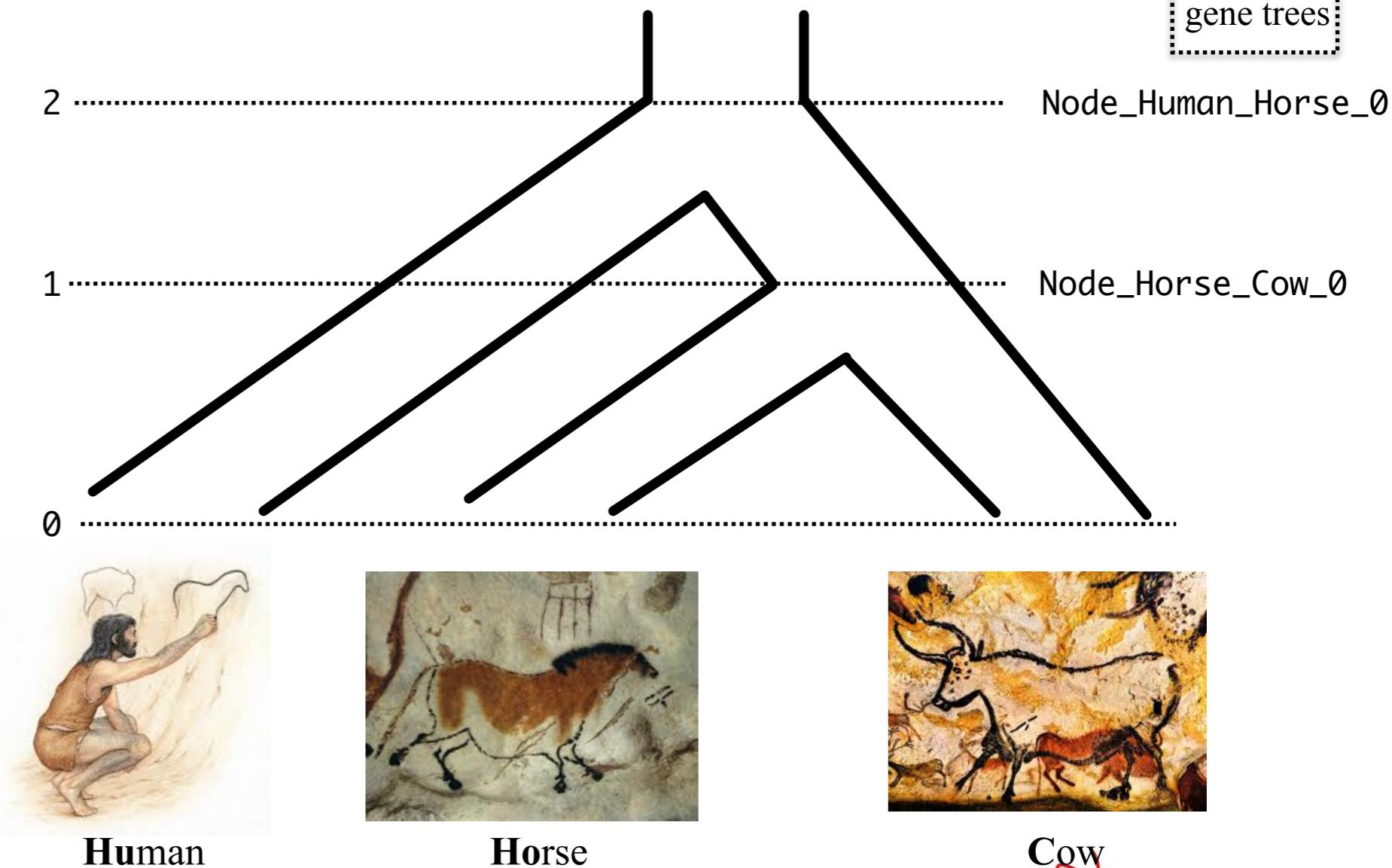
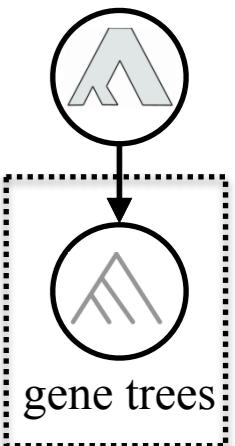


Zuckerkandl & Pauling 1965

Hemoglobin/

```
$ ./alerax --families Zuckerkandl-Pauling.families  
--species-tree data/HuHoCo.sptree  
--prefix Zuckerkandl-Pauling --gene-tree-samples 10  
--rec-model UndatedDL --model-parametrization GLOBAL
```

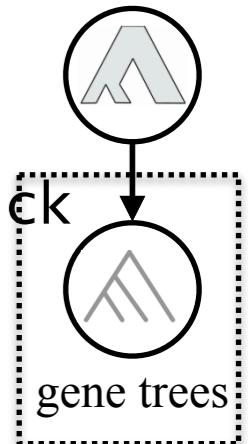
DL
species tree



Hemoglobin/

```
$ ./alerax --families Zuckerkandl-Pauling.families  
--species-tree data/HuHoCo.sptree  
--prefix Zuckerkandl-Pauling --gene-tree-samples 10  
--rec-model UndatedDL --model-parametrization GLOBAL
```

DL
species tree



species tree: Zuckerkandl-Pauling/species_trees/starting_species_tree.newick
(Human:2.000000,(Horse:1.000000,Cow:1.000000)Node_Horse_Cow_0:1.000000)Node_Human_Horse_0:1.000000;

ML log-likelihood (DL model): Zuckerkanld-Pauling/per_family_likelihoods.txt
Zuckerkandl-Pauling -14.1803

ML rates (DL model): Zuckerkandl-Pauling/model_parameters/model_parameters.txt

Human 0.306444 0.240822

10 random reconciled gene trees **sampled according to joint likelihood**

Zuckerkandl-Pauling/reconciliations/all/Zuckerkandl-Pauling.rec_uml

```
((Human_beta:1,Human_delta:1).D@Human:2,Horse_beta-delta:1).Node_Human_Horse_0:2,(Human_gamma:3,(Cow_gamma:2,  
Cow_beta-delta:2).D@Node_Horse_Cow_0:1).Node_Human_Horse_0:2).D@Node_Human_Horse_0;  
(((Human_beta:1,Human_delta:1).D@Human:2,Horse_beta-delta:1).Node_Human_Horse_0:2,(Human_gamma:3,(Cow_gamma:2,  
Cow_beta-delta:2).Node_Horse_Cow_0.D@Cow:1).Node_Human_Horse_0:2).D@Node_Human_Horse_0;  
((Horse_beta-delta:2,((Human_beta:1,Human_delta:1).Node_Human_Horse_0.D@Human:1,(Human_gamma:3,(Cow_gamma:2,  
Cow_beta-delta:2).Node_Horse_Cow_0.D@Cow:1).Node_Human_Horse_0:2).D@Node_Human_Horse_0:2).D@Node_Human_Horse_0;  
(((Human_beta:1,Human_delta:1).D@Human:2,Horse_beta-delta:1).Node_Human_Horse_0:2,(Human_gamma:3,(Cow_gamma:2,  
Cow_beta-delta:2).Node_Horse_Cow_0.D@Cow:1).Node_Human_Horse_0:2).D@Node_Human_Horse_0;  
(((Human_beta:1,Human_delta:1).D@Human:2,Horse_beta-delta:1).Node_Human_Horse_0:2,(Human_gamma:3,(Cow_gamma:2,  
Cow_beta-delta:2).Node_Horse_Cow_0.D@Cow:1).Node_Human_Horse_0:2).D@Node_Human_Horse_0;
```

mean number of events per branch in sampled reconciliations:

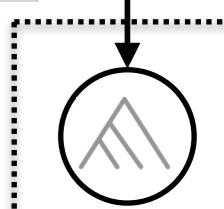
Zuckerkanld-Pauling/reconciliations/perspecies_eventcount.txt

species_label,	speciations,	duplications,	losses,	transfers,	presence,	origination,	copies,	singletons
Cow,	2,	0.9,	1,	0,	1,	0,	2,	2
Horse,	1,	0,	1.1,	0,	1,	0,	1,	1
Human,	3,	1,	0.1,	0,	1,	0,	3,	3
Node_Horse_Cow_0,	2.1,	0.1,	0.1,	0,	1,	0,	2.1,	2.1
Node_Human_Horse_0,	2.1,	1.1,	0,	0,	1,	1,	2.1,	0.2

Hemoglobin/

DL

species tree



gene trees

```
../alerax -s HuHoCo.sptree -f Zukerkandl-Pauling.families  
--rec-model UndatedDL [...]
```

ML log-likelihood (DL model):
Zukerkandl-Pauling -14.1803

mean number of events per branch:

species_label,	speciations,	duplications,	losses,
Cow,	2,	0.9,	1,
Horse,	1,	0,	1.1,
Human,	3,	1,	0.1,
Node_Horse_Cow_0,	2.1,	0.1,	0.1,
Node_Human_Horse_0,	2.1,	1.1,	0,

```
../alerax -s HuHoCo.sptree -f Hemoglobin.tree.ale  
--rec-model UndatedDL [...]
```

ML log-likelihood (DL model):
Hemoglobin -8.05492

mean number of events per branch:

species_label,	speciations,	duplications,	losses
Cow,	2,	1,	0,
Horse,	1,	0,	0,
Human,	3,	2,	0,
Node_Horse_Cow_0,	1,	0,	0,
Node_Human_Horse_0,	1,	0,	0,

ML rates (DL model):
Human 0.306444 0.240822

ML rates (DL model):
Human 0.375001 1e-10

```

Hemoglobin/
$ ./alerax --families Zuckerkandl-Pauling.families
--species-tree data/HuHoCo.sptree
--prefix Zuckerkandl-Pauling-fixed --gene-tree-samples 10
--rec-model UndatedDL --model-parametrization GLOBAL
--fix-rates --d 0.01 --l 0.01
species tree: Zuckerkandl-Pauling-fixed/species_trees/starting_species_tree.newick
(Human:2.000000,(Horse:1.000000,Cow:1.000000)Node_Horse_Cow_0:1.000000)Node_Human_Horse_0:1.000000;
ML log-likelihood (DL model): Zuckerkandl-Pauling-fixed/per_family_likelihoods.txt
Zuckerkandl-Pauling -24.9127

```

ML rates (DL model): Zuckerkandl-Pauling-fixed/model_parameters/model_parameters.txt
Human 0.01 0.01

10 random reconciled gene trees **sampled according to joint likelihood**

Zuckerkandl-Pauling-fixed/reconciliations/all/Zuckerkandl-Pauling.rec_uml

```

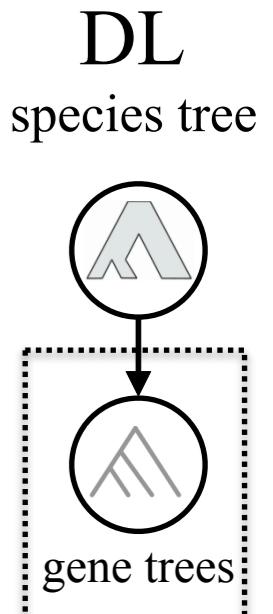
(((Human_beta:1,Human_delta:1).D@Human:2,Horse_beta-delta:1).Node_Human_Horse_0:2,(Human_gamma:3,(Cow_gamma:2,
Cow_beta-delta:2).Node_Horse_Cow_0.D@Cow:1).Node_Human_Horse_0:2).D@Node_Human_Horse_0;

```

mean number of events per branch in sampled reconciliations:

Zuckerkandl-Pauling-fixed/reconciliations/perspecies_eventcount.txt

species_label	speciations	duplications	losses	transfers	presence	origination	copies	singletons
Cow,	2,	1,	1,	0,	1,	0,	2,	2
Horse,	1,	0,	1,	0,	1,	0,	1,	1
Human,	3,	1,	0,	0,	1,	0,	3,	3
Node_Horse_Cow_0,	2,	0,	0,	0,	1,	0,	2,	2
Node_Human_Horse_0,	2,	1,	0,	0,	1,	1,	2,	0



Hemoglobin/

```
$ ./alerax --families Zuckerkandl-Pauling.families  
--species-tree data/HuHoCo.sptree  
--prefix Zuckerkandl-Pauling-fixed --gene-tree-samples 10  
--rec-model UndatedDL --model-parametrization GLOBAL  
--fix-rates --d 0.01 --l 0.01
```

ML log-likelihood (DL model):

Zuckerkandl-Pauling -24.9127

mean number of events per branch:

species_label,	speciations,	duplications,	losses
Cow,	2,	1,	1,
Horse,	1,	0,	1,
Human,	3,	1,	0,
Node_Horse_Cow_0,	2,	0,	0,
Node_Human_Horse_0,	2,	1,	0,

ML rates (DL model):

Human 0.01 0.01

```
./alerax -s HuHoCo.sptree  
-f Zuckerkandl-Pauling.families  
--rec-model UndatedDL [...]
```

ML log-likelihood (DL model):

Zuckerkandl-Pauling -14.1803

mean number of events per branch:

species_label,	speciations,	duplications,	losses,
Cow,	2,	0.9,	1,
Horse,	1,	0,	1.1,
Human,	3,	1,	0.1,
Node_Horse_Cow_0,	2.1,	0.1,	0.1,
Node_Human_Horse_0,	2.1,	1.1,	0,

ML rates (DL model):

Human 0.306444 0.240822

```
./alerax -s HuHoCo.sptree  
-f Hemoglobin.tree.ale  
--rec-model UndatedDL [...]
```

ML log-likelihood (DL model):

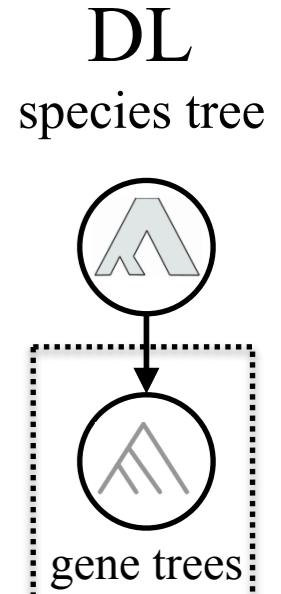
Hemoglobin -8.05492

mean number of events per branch:

species_label,	speciations,	duplications,	losses
Cow,	2,	1,	0,
Horse,	1,	0,	0,
Human,	3,	2,	0,
Node_Horse_Cow_0,	1,	0,	0,
Node_Human_Horse_0,	1,	0,	0,

ML rates (DL model):

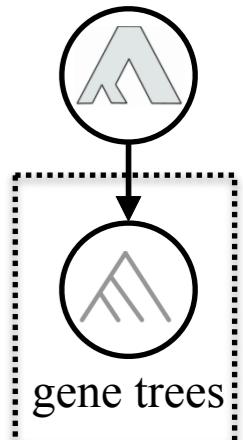
Human 0.375001 1e-10



Hemoglobin/Zuckerkandl-Pauling-fixed

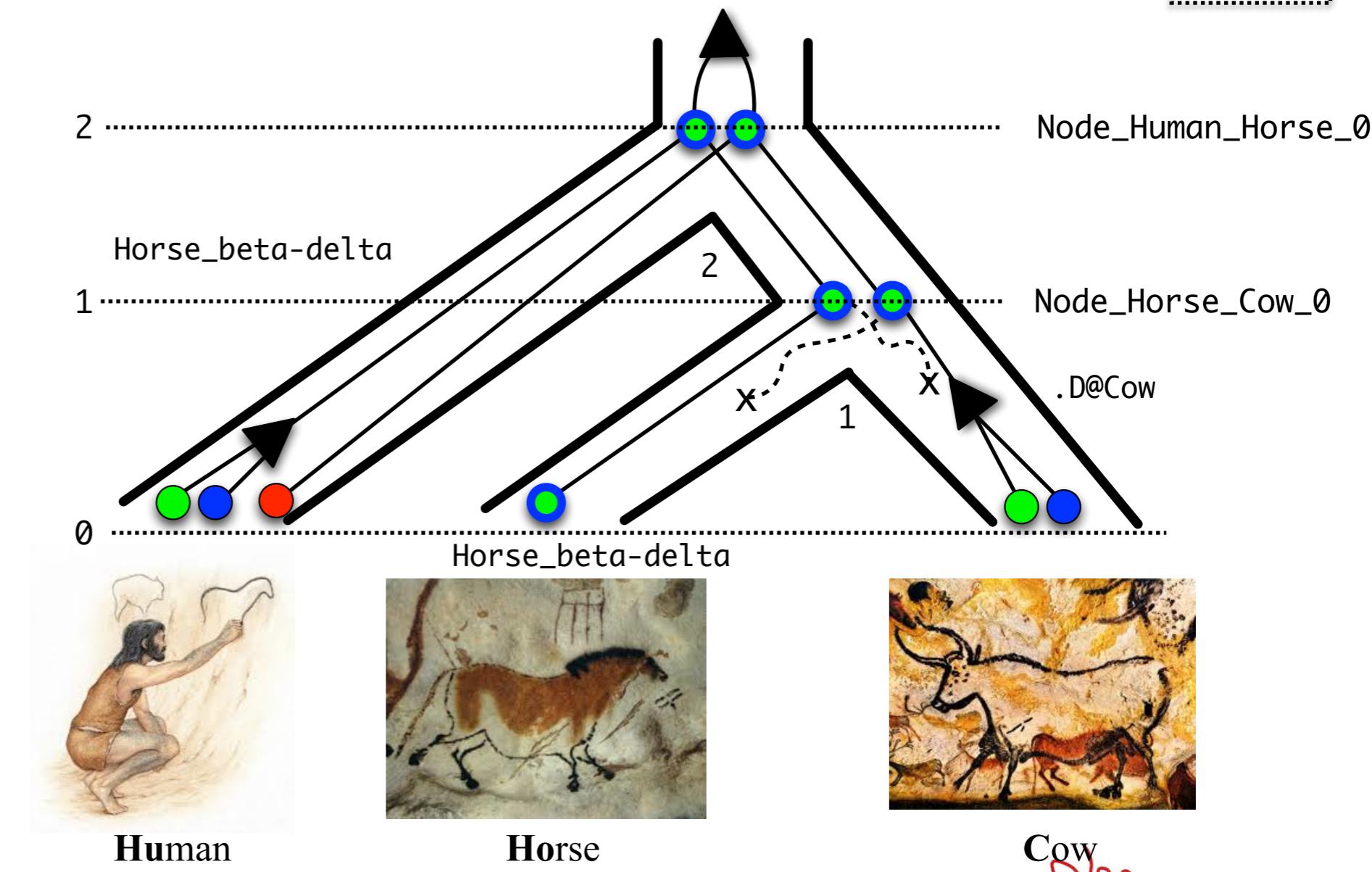
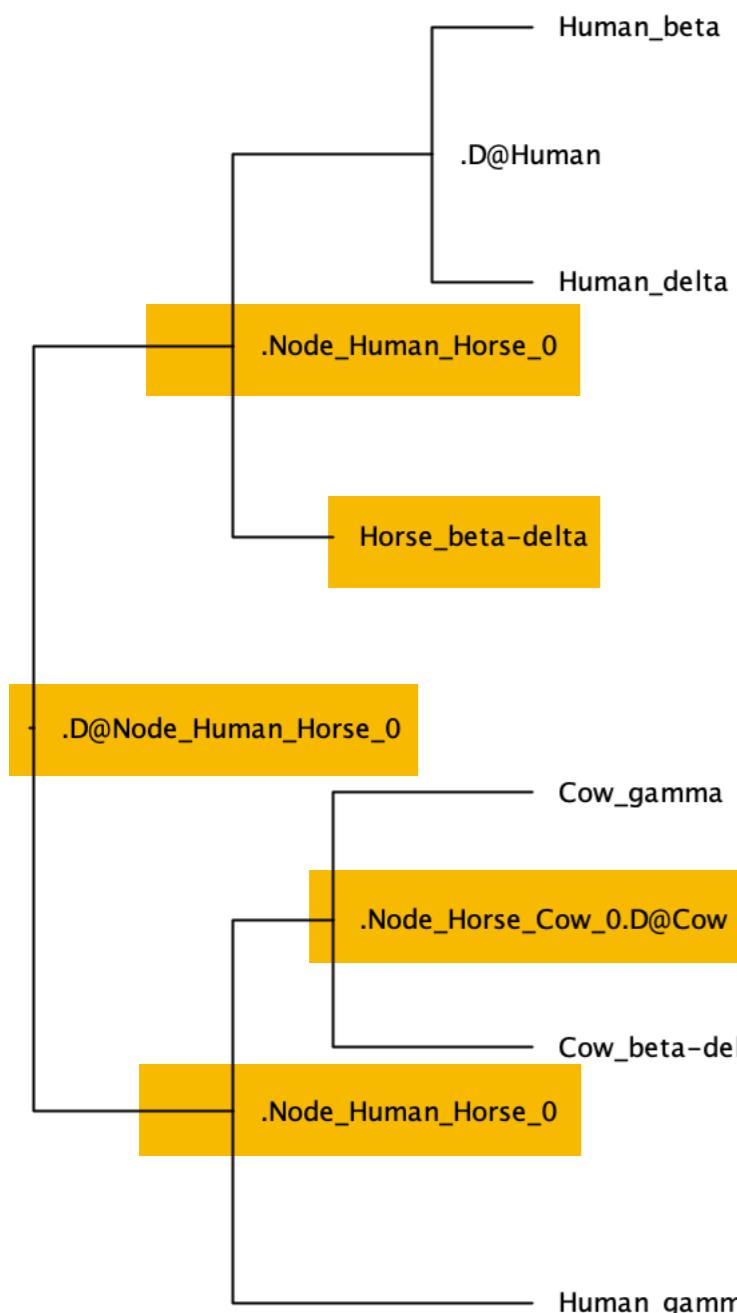
((Human_beta:1, Human_delta:1).D@Human:2, Horse_beta-delta:1).Node_Human_Horse_0:2, (Human_gamma:3, (Cow_gamma:2, Cow_beta-delta:2).Node_Horse_Cow_0.D@Cow:1).Node_Human_Horse_0:2).D@Node_Human_Horse_0;

DL
species tree



mean number of events per branch in sampled reconciliations:

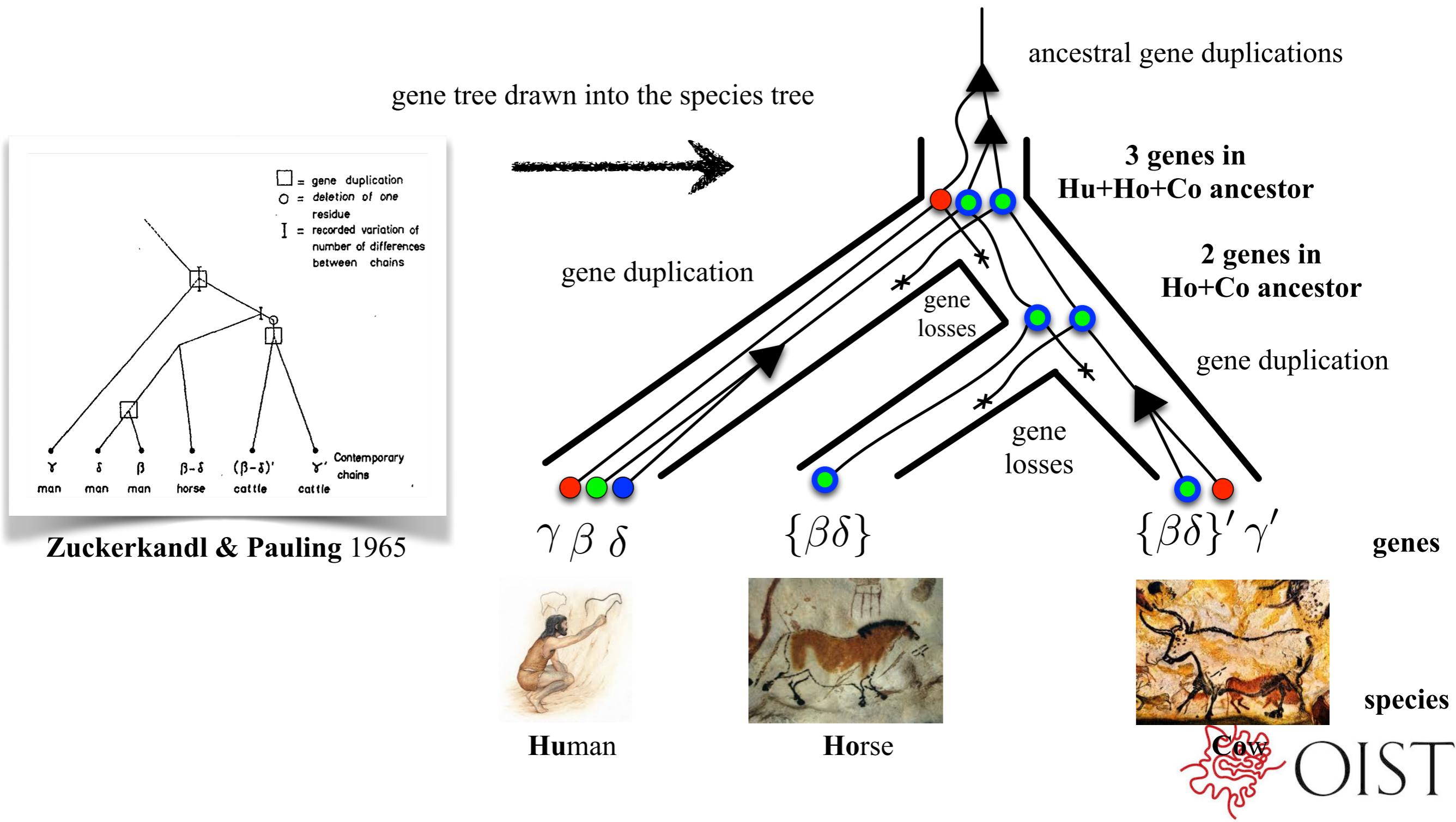
species_label,	speciations,	duplications,	losses,	transfers,	presence,	origination,	copies,	singletons
Cow,	2,	1,	1,	0,	1,	0,	2,	2
Horse,	1,	0,	1,	0,	1,	0,	1,	1
Human,	3,	1,	0,	0,	1,	0,	3,	3
Node_Horse_Cow_0,	2,	0,	0,	0,	1,	0,	2,	2
Node_Human_Horse_0,	2,	1,	0,	0,	1,	1,	2,	0



Hemoglobin/

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.



Hemoglobin/

```
$ ./alerax --families Zuckerkandl-Pauling_OUT.families
```

```
--species-tree data/HuHoCo_OUT.sptree
```

```
--prefix Zuckerkandl-Pauling-OUT-fixed --gene-tree-samples 10
```

```
--rec-model UndatedDL --model-parametrization GLOBAL
```

```
--fix-rates --d 0.01 --l 0.01
```

species tree: Zuckerkandl-Pauling-OUT-fixed/species_trees/startng_species_tree.newick

```
((Human:2.000000,(Horse:1.000000,Cow:1.000000)Node_Horse_Cow_0:1.000000)Node_Human_Horse_0:1.000000,OUT:4.000000)Node_Human_OUT_0:1.000000;
```

ML log-likelihood (DL model): Zuckerkandl-Pauling-OUT-fixed/per_family_likelihoods.txt

Zuckerkandl-Pauling -39.1543

ML rates (DL model): Zuckerkandl-Pauling-OUT-fixed/model_parameters/model_parameters.txt

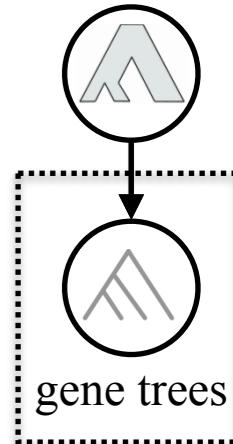
Human 0.01 0.01

10 random reconciled gene trees **sampled according to joint likelihood**

Zuckerkandl-Pauling-OUT-fixed/reconciliations/all/Zuckerkandl-Pauling.rec.uml

```
((Human_gamma:4,(((Human_beta:1,Human_delta:1).D@Human:1,Horse_beta-delta:2).Node_Human_Horse_0:1,(Cow_beta-delta:1,Cow_gamma:1).Node_Human_Horse_0.Node_Horse_Cow_0.D@Cow:2).D@Node_Human_Horse_0:1).D@Node_Human_Horse_0:6,OUT_OUT:6).Node_Human_OUT_0;  
((Human_gamma:4,(((Human_beta:1,Human_delta:1).D@Human:1,Horse_beta-delta:2).Node_Human_Horse_0:1,(Cow_beta-delta:1,Cow_gamma:1).Node_Human_Horse_0.Node_Horse_Cow_0.D@Cow:2).D@Node_Human_Horse_0:1).D@Node_Human_Horse_0:6,OUT_OUT:6).Node_Human_OUT_0;  
((Human_gamma:4,(((Human_beta:1,Human_delta:1).D@Human:1,Horse_beta-delta:2).Node_Human_Horse_0:1,(Cow_beta-delta:1,Cow_gamma:1).Node_Human_Horse_0.Node_Horse_Cow_0.D@Cow:2).D@Node_Human_Horse_0:1).D@Node_Human_Horse_0:6,OUT_OUT:6).Node_Human_OUT_0;  
((Human_gamma:4,(((Human_beta:1,Human_delta:1).D@Human:1,Horse_beta-delta:2).Node_Human_Horse_0:1,(Cow_beta-delta:1,Cow_gamma:1).Node_Human_Horse_0.Node_Horse_Cow_0.D@Cow:2).D@Node_Human_Horse_0:1).D@Node_Human_Horse_0:6,OUT_OUT:6).Node_Human_OUT_0;  
((Human_gamma:4,(((Human_beta:1,Human_delta:1).D@Human:1,Horse_beta-delta:2).Node_Human_Horse_0:1,(Cow_beta-delta:1,Cow_gamma:1).Node_Human_Horse_0.Node_Horse_Cow_0.D@Cow:2).D@Node_Human_Horse_0:1).D@Node_Human_Horse_0:6,OUT_OUT:6).Node_Human_OUT_0;  
((Human_gamma:4,(((Human_beta:1,Human_delta:1).D@Human:1,Horse_beta-delta:2).Node_Human_Horse_0:1,(Cow_beta-delta:1,Cow_gamma:1).Node_Human_Horse_0.Node_Horse_Cow_0.D@Cow:2).D@Node_Human_Horse_0:1).D@Node_Human_Horse_0:6,OUT_OUT:6).Node_Human_OUT_0;  
((Human_gamma:4,(((Human_beta:1,Human_delta:1).D@Human:1,Horse_beta-delta:2).Node_Human_Horse_0:1,(Cow_beta-delta:1,Cow_gamma:1).Node_Human_Horse_0.Node_Horse_Cow_0.D@Cow:2).D@Node_Human_Horse_0:1).D@Node_Human_Horse_0:6,OUT_OUT:6).Node_Human_OUT_0;  
((Human_gamma:4,(((Human_beta:1,Human_delta:1).D@Human:1,Horse_beta-delta:2).Node_Human_Horse_0:1,(Cow_beta-delta:1,Cow_gamma:1).Node_Human_Horse_0.Node_Horse_Cow_0.D@Cow:2).D@Node_Human_Horse_0:1).D@Node_Human_Horse_0:6,OUT_OUT:6).Node_Human_OUT_0;  
((Human_gamma:4,(((Human_beta:1,Human_delta:1).D@Human:1,Horse_beta-delta:2).Node_Human_Horse_0:1,(Cow_beta-delta:1,Cow_gamma:1).Node_Human_Horse_0.Node_Horse_Cow_0.D@Cow:2).D@Node_Human_Horse_0:1).D@Node_Human_Horse_0:6,OUT_OUT:6).Node_Human_OUT_0;  
((Human_gamma:4,(((Human_beta:1,Human_delta:1).D@Human:1,Horse_beta-delta:2).Node_Human_Horse_0:1,(Cow_beta-delta:1,Cow_gamma:1).Node_Human_Horse_0.Node_Horse_Cow_0.D@Cow:2).D@Node_Human_Horse_0:1).D@Node_Human_Horse_0:6,OUT_OUT:6).Node_Human_OUT_0;
```

DL
species tree



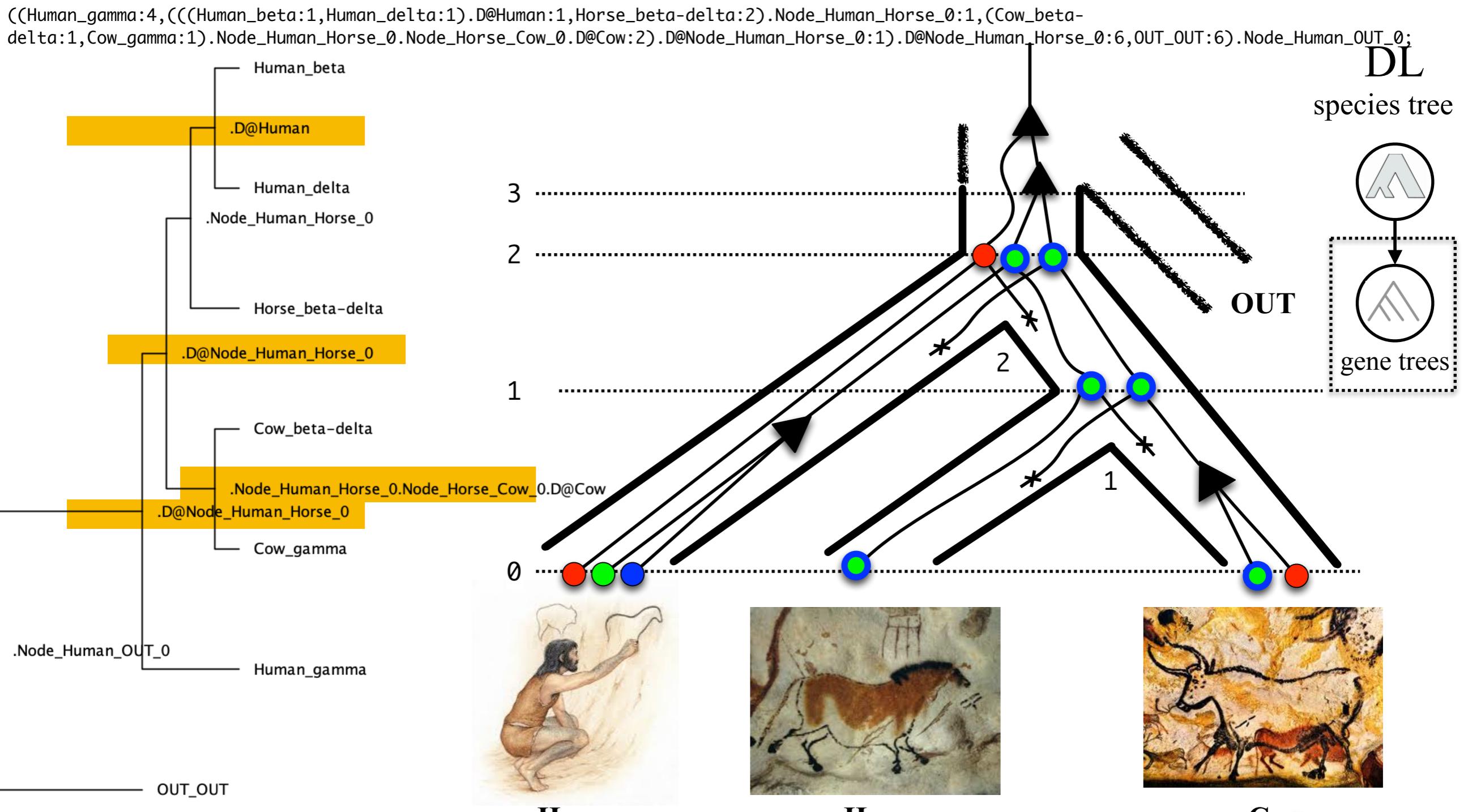
mean number of events per branch in sampled reconciliations: perspecies_eventcount.txt

species_label, speciations, duplications, losses, transfers, presence, origination, copies, singletons

Cow,	2,	1,	1,	0,	1,	0,	2,	2
Horse,	1,	0,	1,	0,	1,	0,	1,	1
Human,	3,	1,	1,	0,	1,	0,	3,	3
Node_Horse_Cow_0,	2,	0,	1,	0,	1,	0,	2,	2
Node_Human_Horse_0,	3,	2,	0,	0,	1,	0,	3,	2
Node_Human_OUT_0,	1,	0,	0,	0,	1,	1,	1,	1
OUT,	1,	0,	0,	0,	1,	0,	1,	1

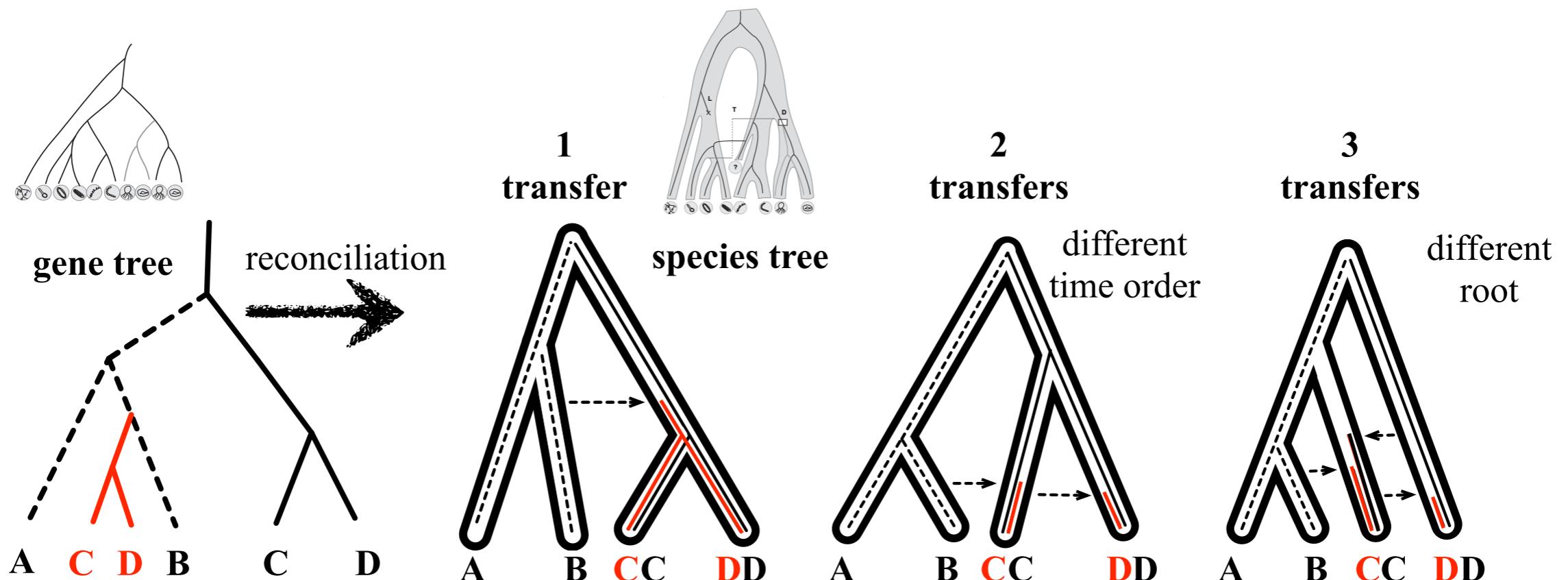
Hemoglobin/

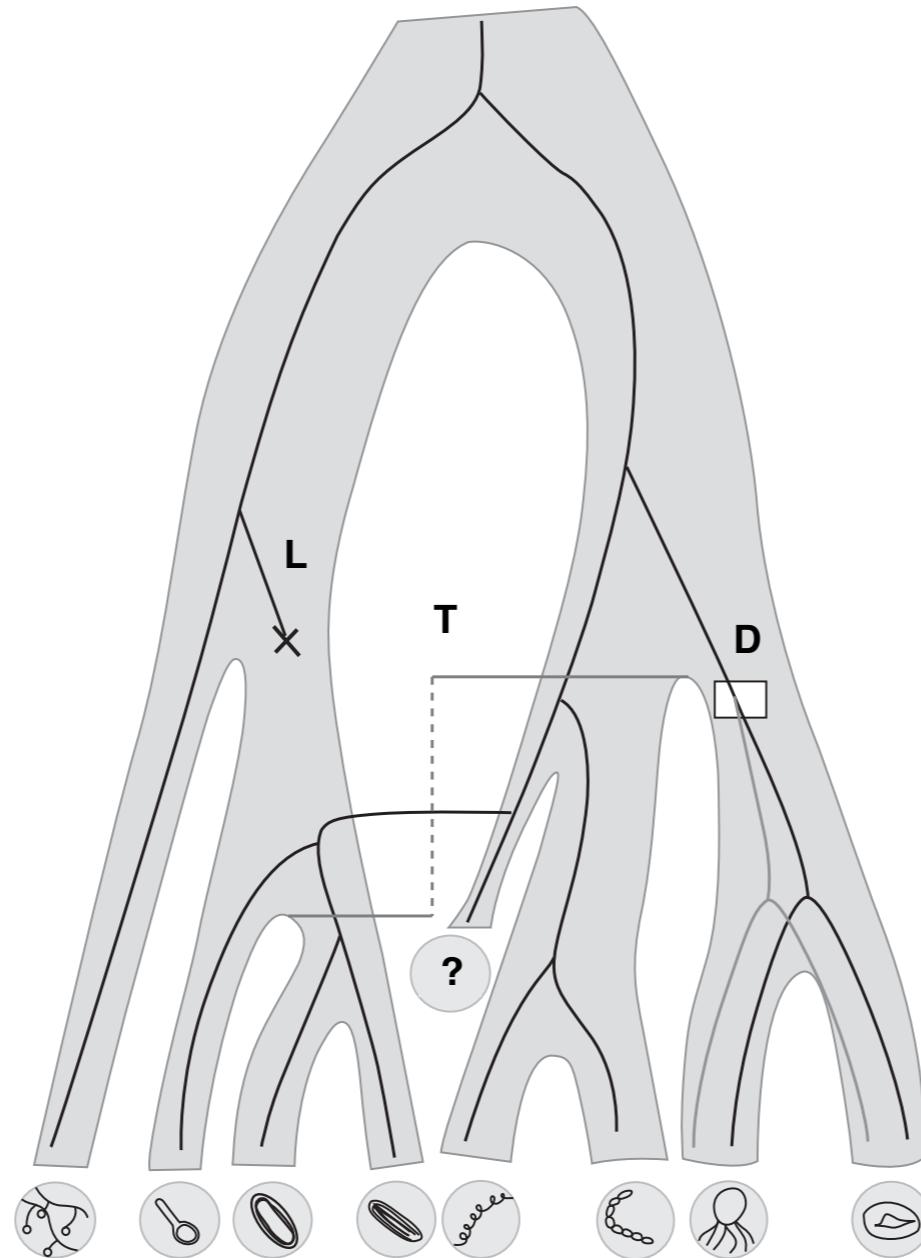
```
$ ./alerax --families Zuckerkandl-Pauling_OUT.families  
--species-tree data/HuHoCo_OUT.sptree [...]  
--fix-rates --d 0.01 --l 0.01
```



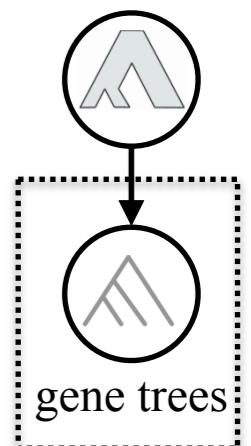
Horizontal gene transfer as information

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



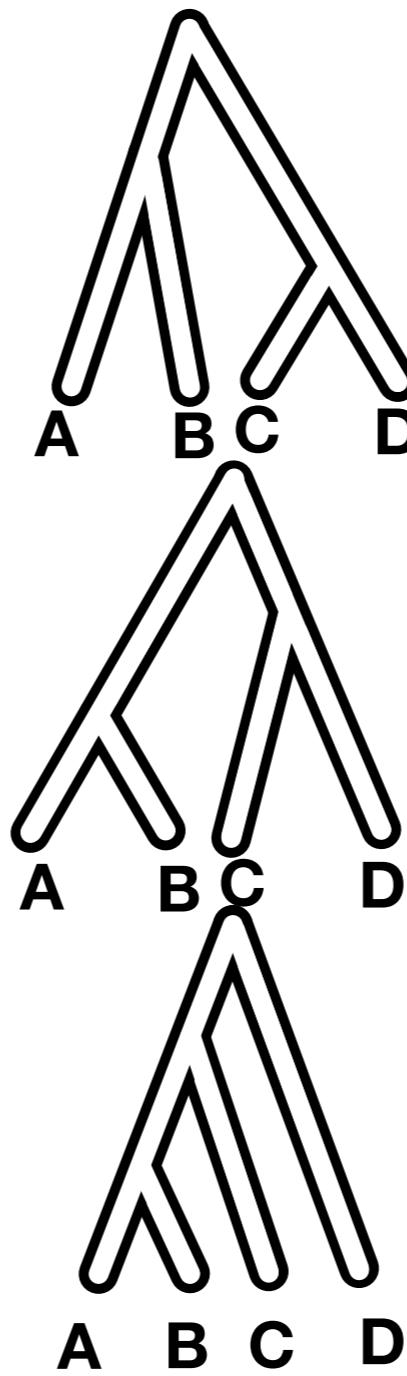
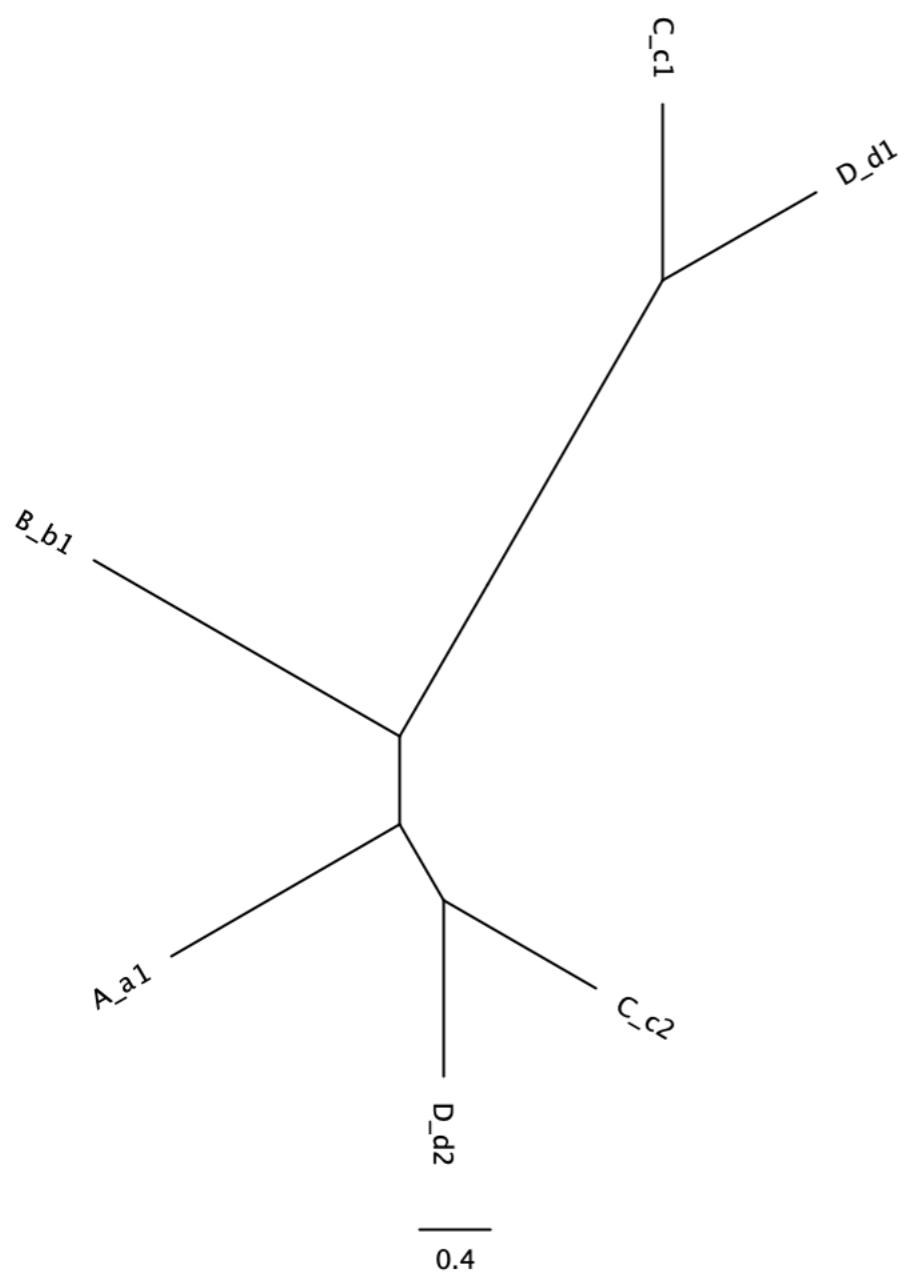


DTL
species tree



abcd/

gene tree
g.tree

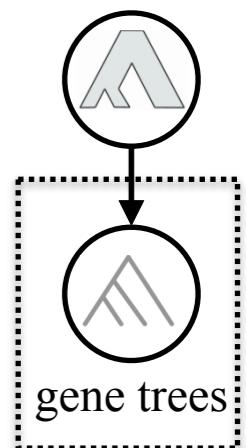


abcd_S.tree

cdab_S.tree

abc-d_S.tree

DTL
species tree



abcd/

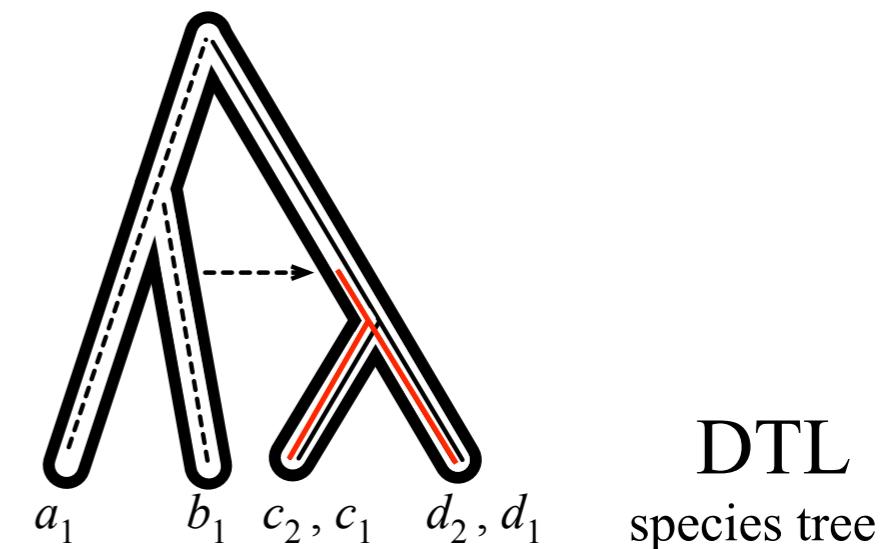
\$ ALEobserve g.tree

\$ ALEml abcd_S.tree g.tree.ale sample=10

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 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 1.2 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)
```

abcd_S.tree_g.tree.ale.ml_rec

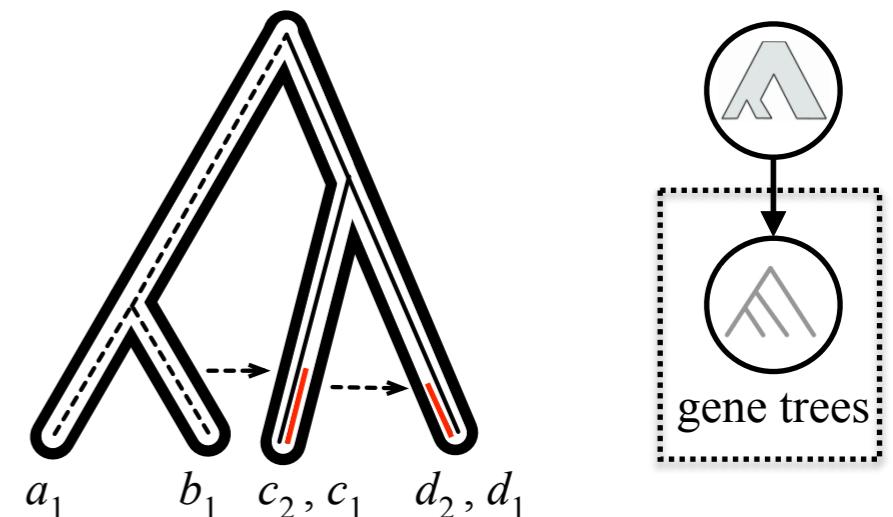


\$ ALEml cdab_S.tree g.tree.ale sample=10

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 0.1 0 1  
S_terminal_branch b 0 0.1 0 1  
S_terminal_branch c 0 1.2 0 2  
S_terminal_branch d 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)
```

cdab_S.tree_g.tree.ale.ml_rec

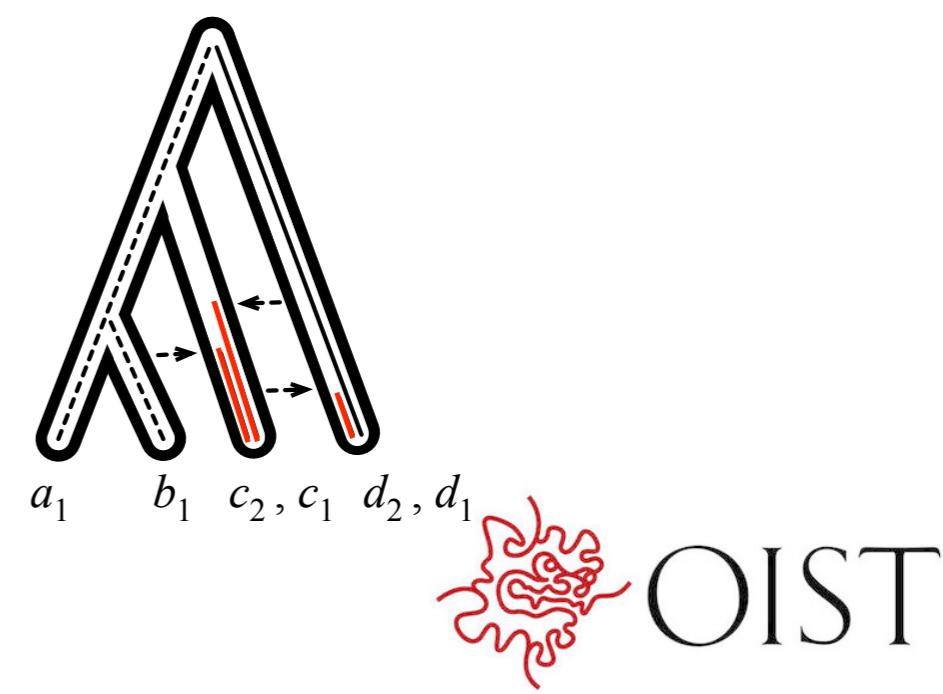


\$ ALEml abc-d_S.tree g.tree.ale sample=10

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 1  
S_terminal_branch b 0 0 0 1  
S_terminal_branch c 0 1 0 2  
S_terminal_branch d 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)
```

abc-d_S.tree_g.tree.ale.ml_rec



abcd/

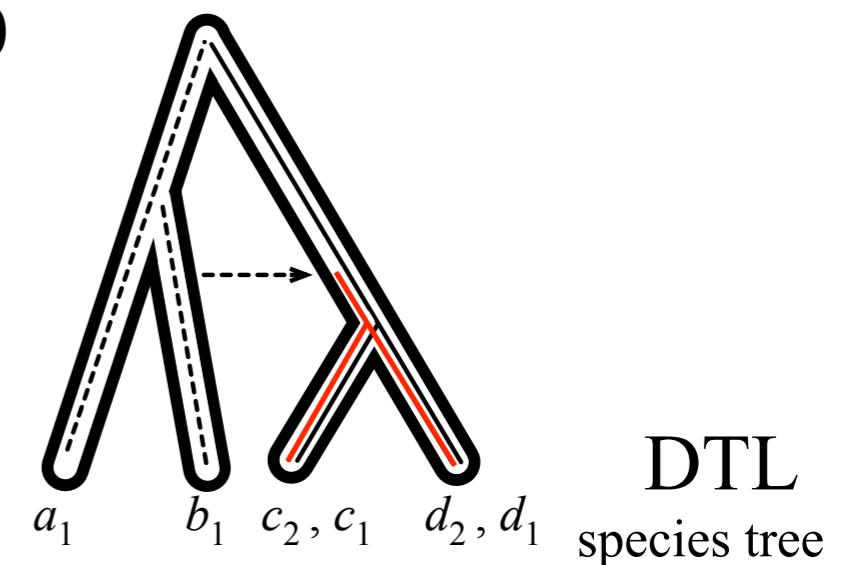
\$ ALEobserve g.tree

\$./ALEml abcd_S.tree g.tree.ale sample=10
delta=0.0 tau=0.01 lambda=0.01

log-likelihood → ..
>logl: -8.61538

events → ..
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)

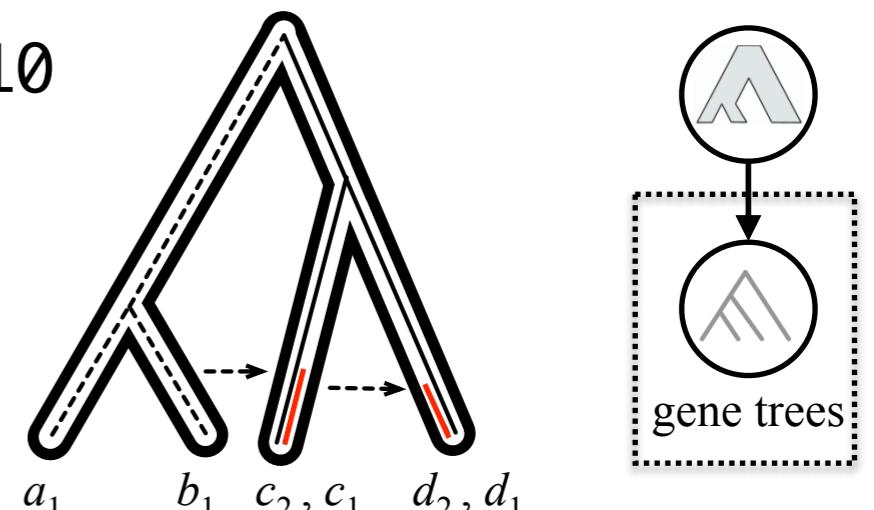


\$./ALEml cdab_S.tree g.tree.ale sample=10
delta=0.0 tau=0.01 lambda=0.01

log-likelihood → ..
>logl: -16.7168

events → ..
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)



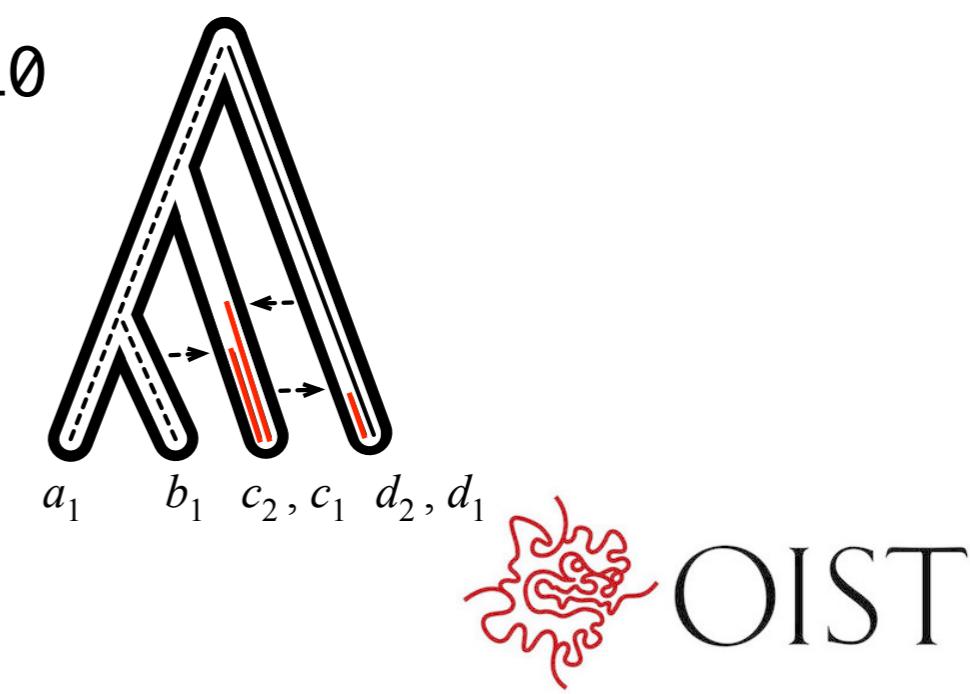
\$./ALEml abc-d_S.tree g.tree.ale sample=10
delta=0.0 tau=0.01 lambda=0.01

log-likelihood → ..
>logl: -16.8563

events → ..
rate of Duplications Transfers Losses
ML 3.74258e-06 1.06978 1e-10

of Duplications Transfers Losses Speciations
Total 0 2 0 3

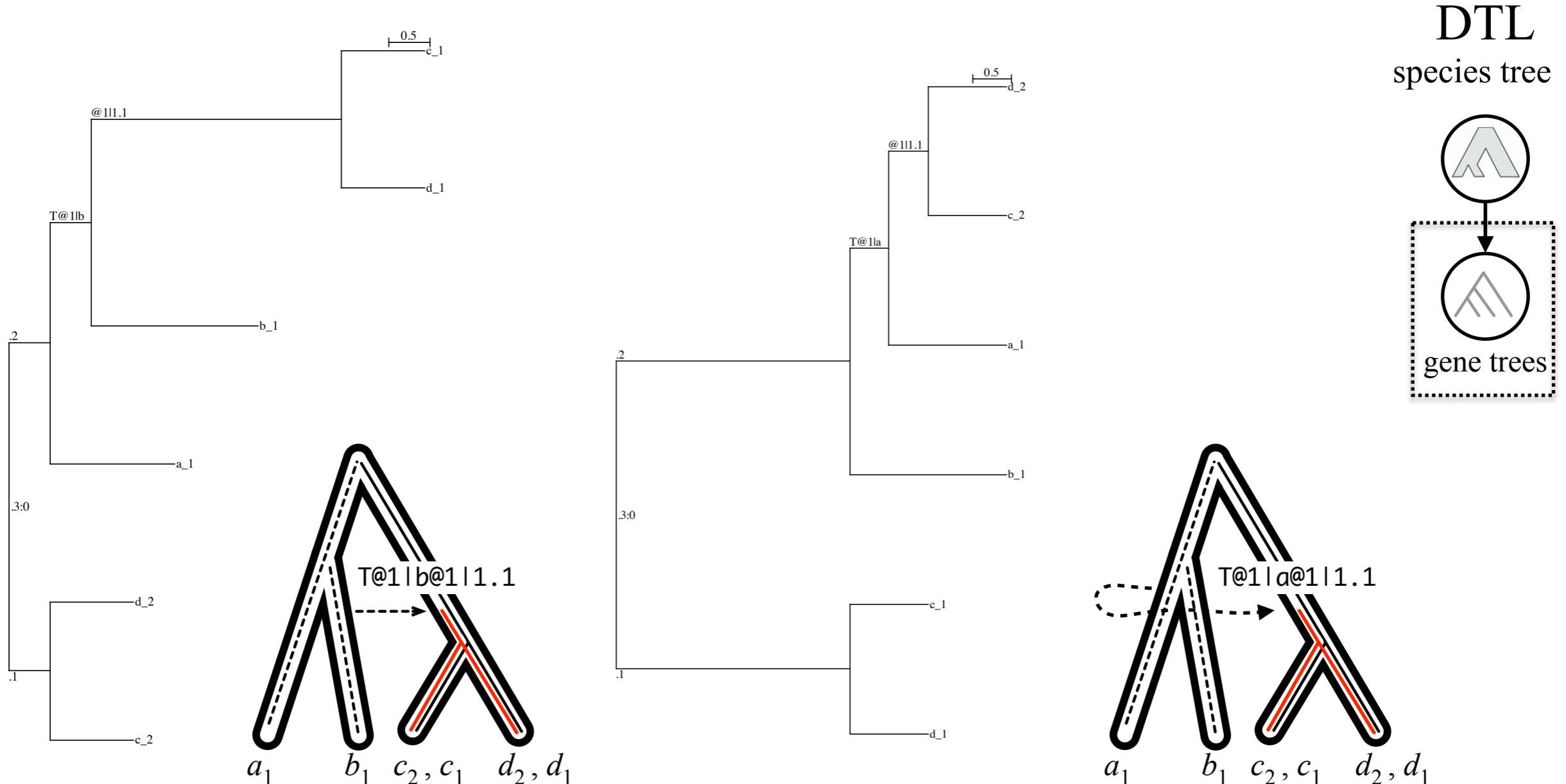
of Duplications Transfers Losses copies
S_terminal_branch a 0 0 0 0 1
S_terminal_branch b 0 0 0 0 1
S_terminal_branch c 0 0 1 0 2
S_terminal_branch d 0 0 1 0 2
S_internal_branch 1 0 0 0 0 1
S_internal_branch 2 0 0 0 0 1
S_internal_branch 3 0 0 0 0 1
abc-d_S.tree_g.tree.ale.ml_rec (END)



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;
```



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

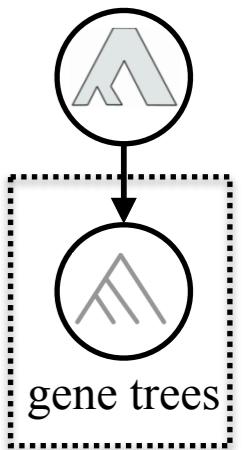
```
((d_1:1,c_1:1).2:3,(b_1:2,(a_1:1.5,(c_2@0|c:1,d_2@0|d:1)T@0|-1:0.5)T@0|a:0.5).1:3).3:0;
((a_1.1:1.5,(c_2:1,d_2:1).2:0.5).3:0.5,(b_1.1:2,(d_1:1,c_1:1).2:3)@3|3.3:0.5)T@3|3:0;
((c_2:1,d_2:1).2:0.5,(a_1:1.5,(b_1:2,(d_1@0|d:1,c_1@0|c:1)T@0|-1:3)T@0|b:0.5).1:0.5).3:0;
((c_2:1,d_2:1).2:0.5,(a_1:1.5,(b_1:2,(d_1@0|d:1,c_1@0|c:1)T@0|-1:3)T@0|b:0.5).1:0.5).3:0;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

DTL

species tree

```
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
```



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

abcd/

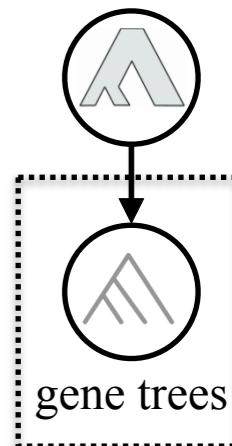
\$./ALEml cdab_S.tree g.tree.ale sample=10 delta=0.0 tau=0.01 lambda=0.01

```
((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_1:1.5,(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;
```

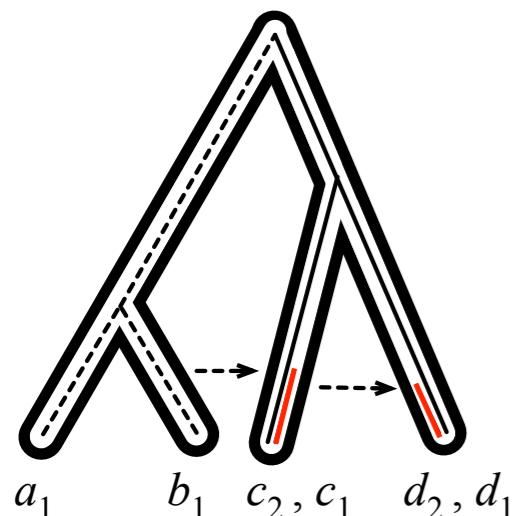
\$./ALEml cdab_S.tree g.tree.ale sample=10 delta=0.01 tau=0.0 lambda=0.01

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

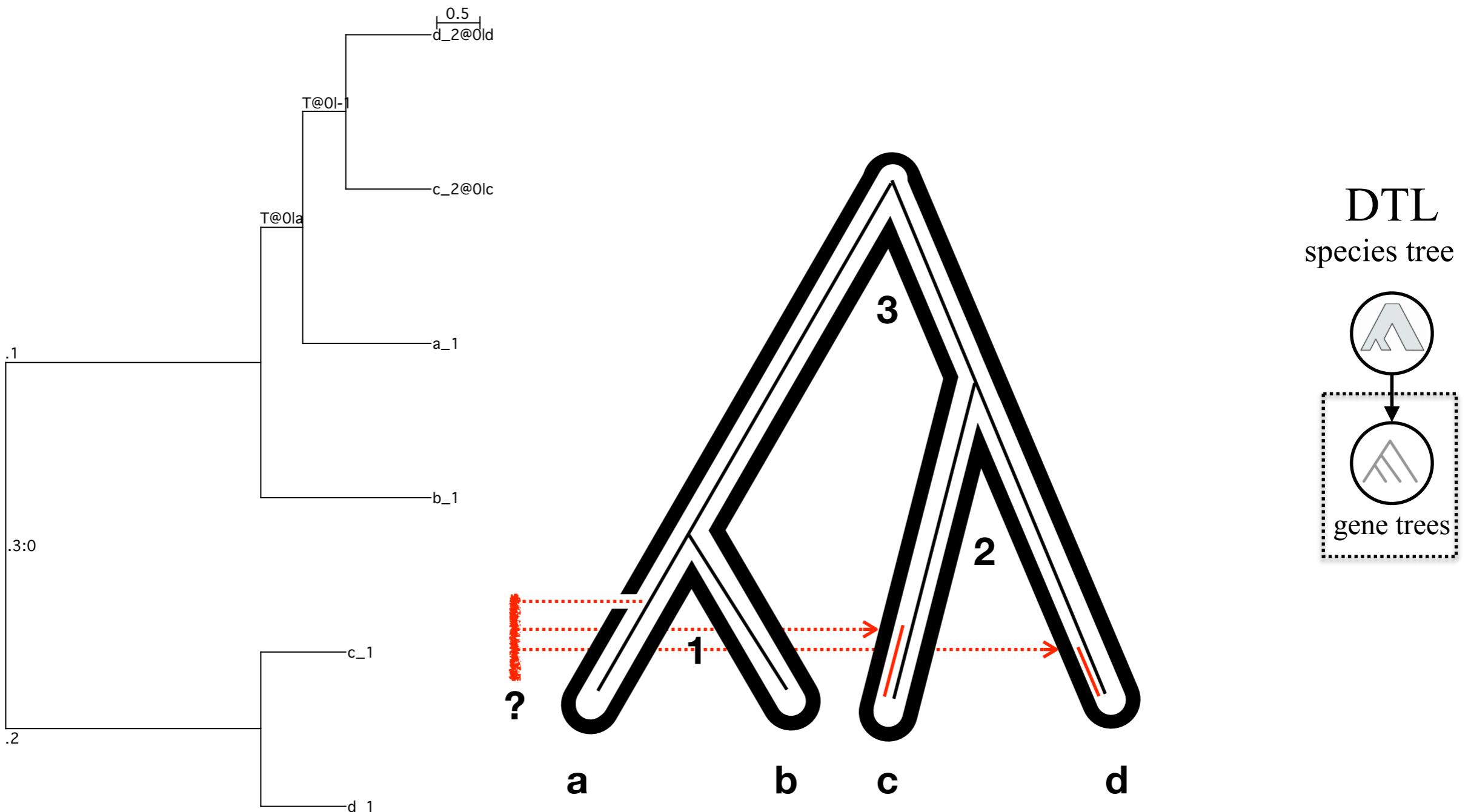


```
((d_1:1,c_1:1).2:3,(b_1:2,(a_1:1.5,(c_2:1,d_2@0|d:1)@0|cT@0|c:0.5)T@0|a:0.5).1:3).3:0;  
((c_2:1,d_2:1).2:0.5,(a_1:1.5,(b_1:2,(d_1@0|d:1,c_1@0|c:1)T@0|-1:3)T@0|b:0.5).1:0.5).3:0;  
((d_1:1,c_1:1).2:3,(b_1:2,(a_1:1.5,(c_2@0|c:1,d_2@0|d:1)T@0|-1:0.5)T@0|a:0.5).1:3).3:0;  
((c_2:1,d_2:1).2:0.5,(a_1:1.5,(b_1:2,(d_1@0|d:1,c_1:1)@0|cT@0|c:3)T@0|b:0.5).1:0.5).3:0;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

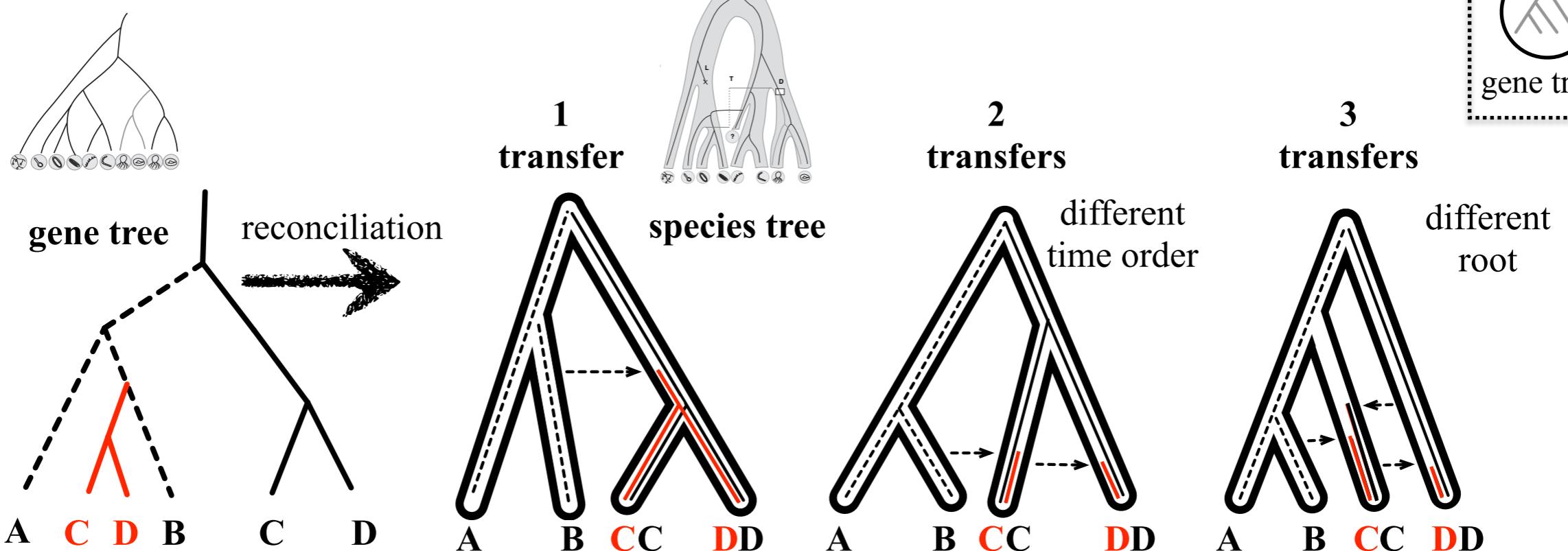
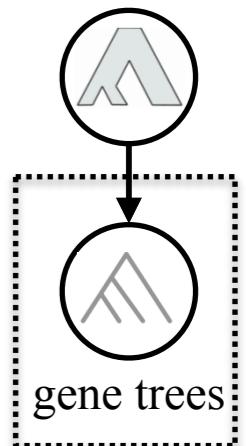


view the tree using either
phylo.io, seaview or FigTree!

abcd/

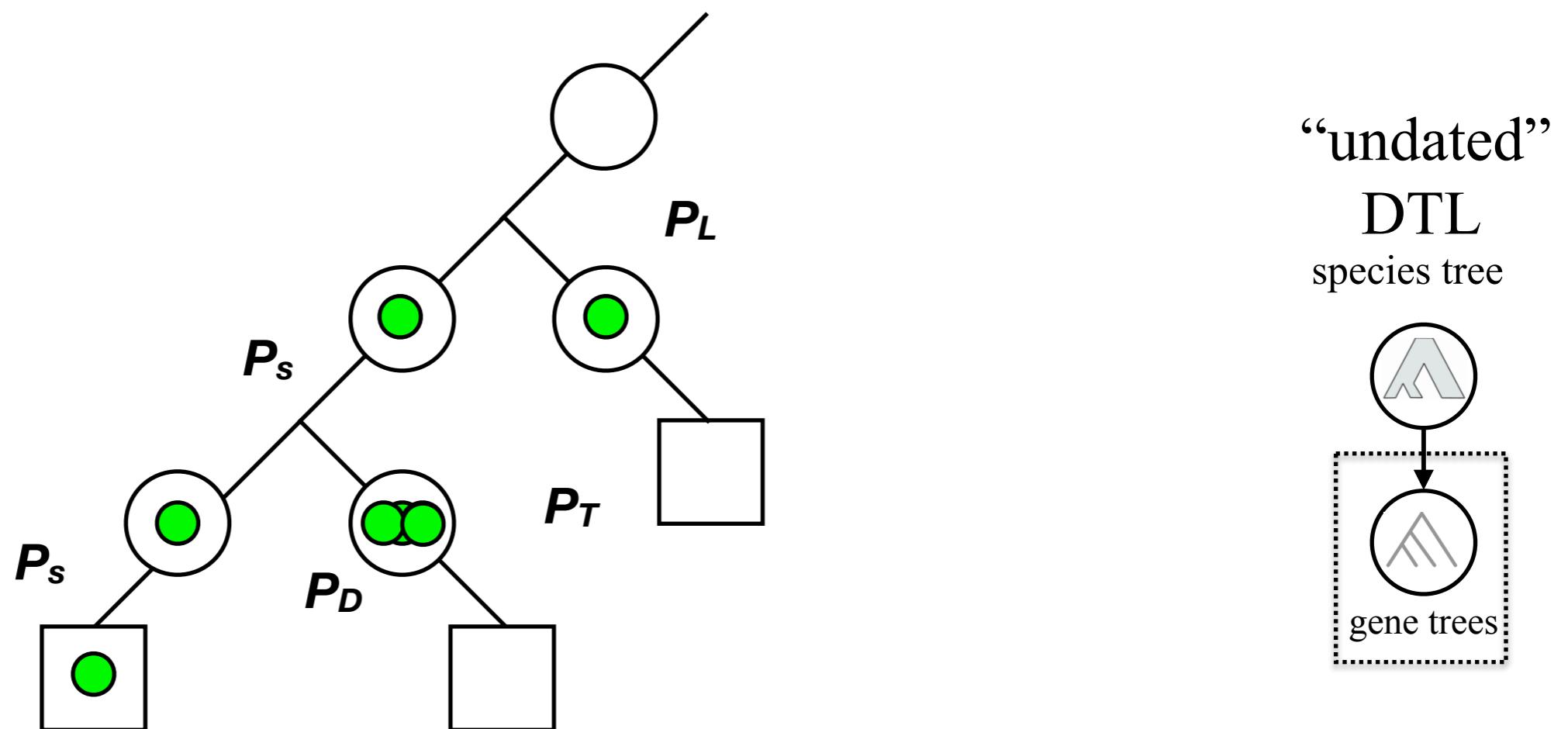
```
$ ./ALEobserve g.tree  
$ ./ALEml abcd_S.tree g.tree.ale sample=10  
$ ./ALEml cdab_S.tree g.tree.ale sample=10  
$ ./ALEml abc-d_S.tree g.tree.ale sample=10
```

DTL
species tree



abcd/

“undated”
DTL



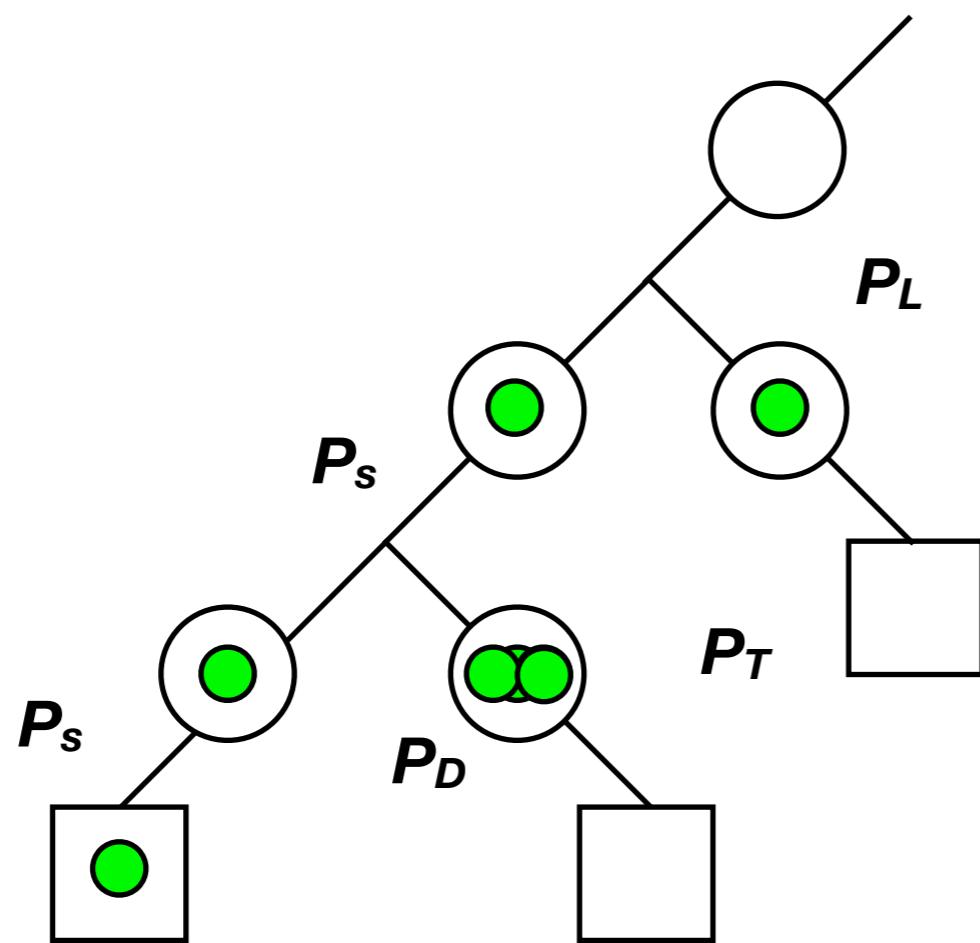
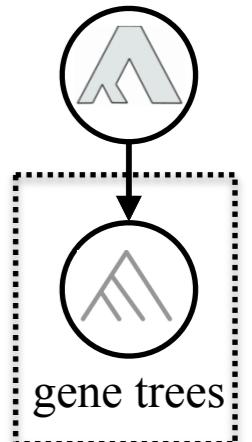
abcd/

“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
```

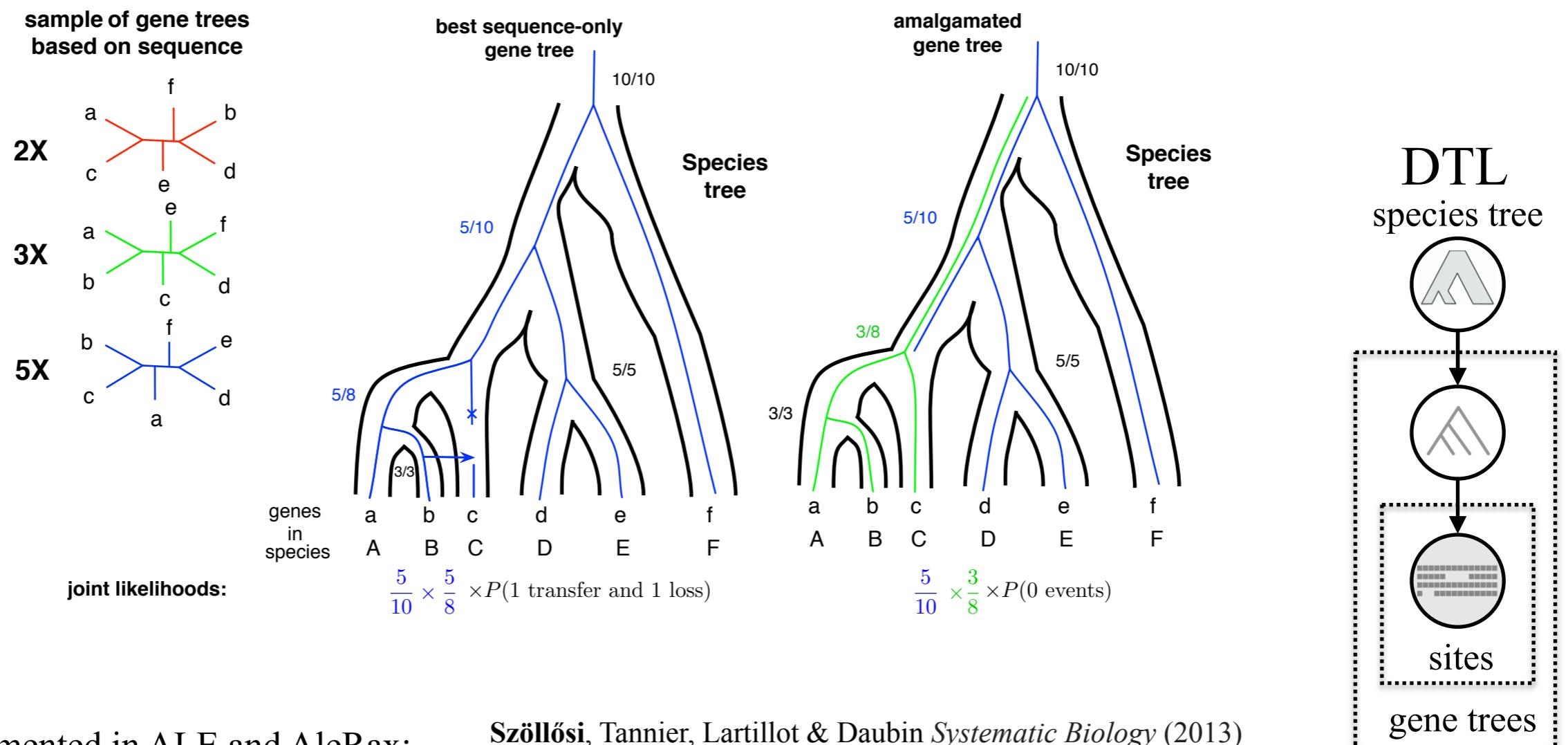
“undated”
DTL
species tree



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 and AleRax:

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

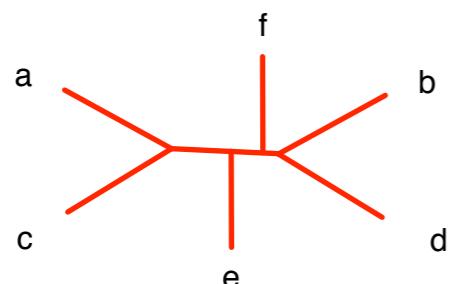
<https://github.com/BenoitMorel/AleRax>

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

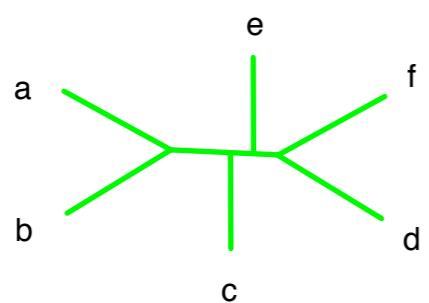
Szöllősi, Rosikiewicz, Boussau, Tannier & Daubin *Systematic Biology* (2013)
Efficient exploration of the space of reconciled gene trees

abcdef/

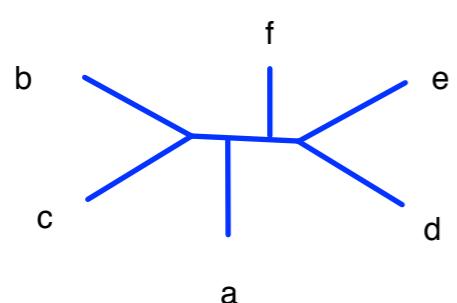
red.tree



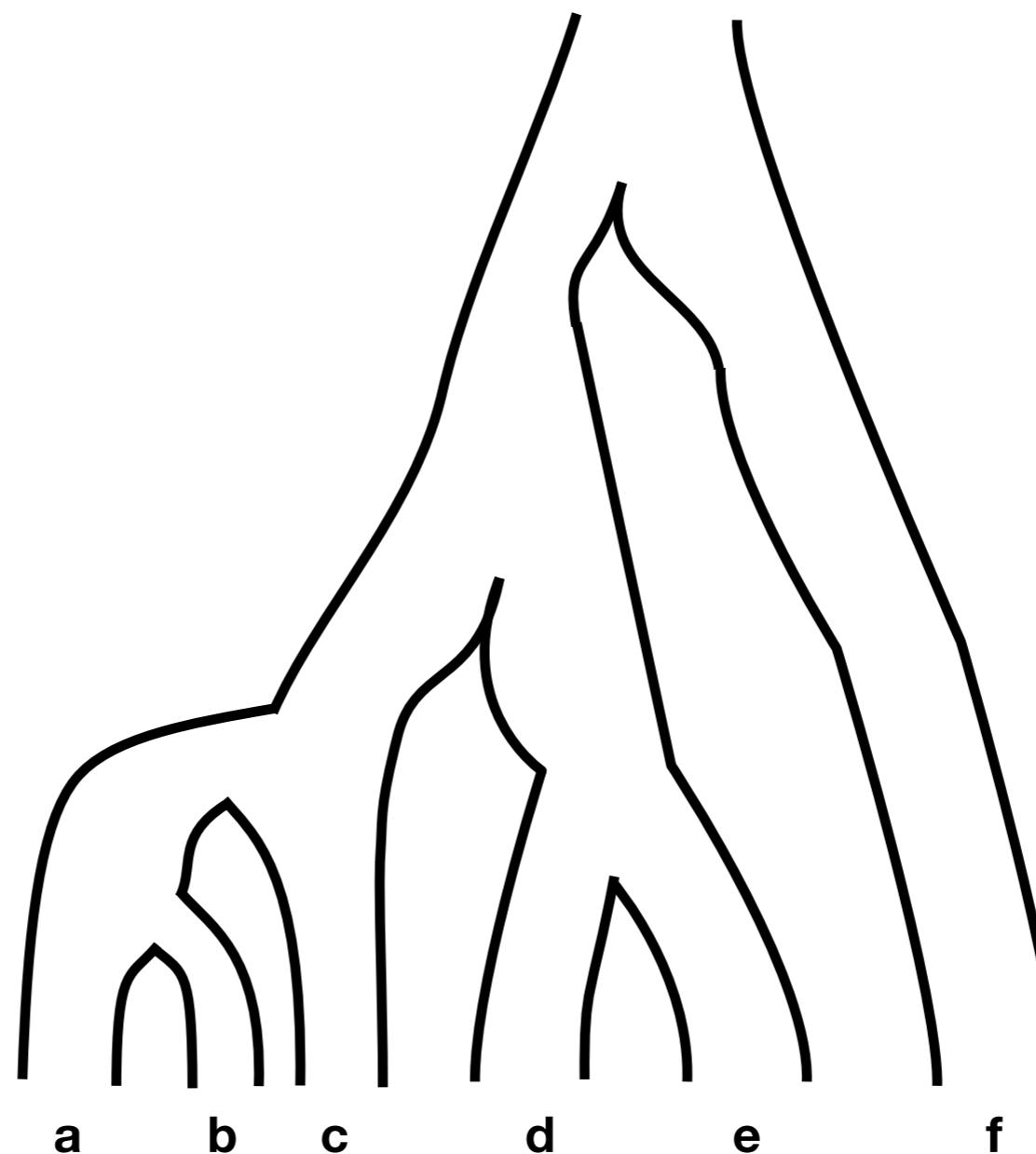
green.tree



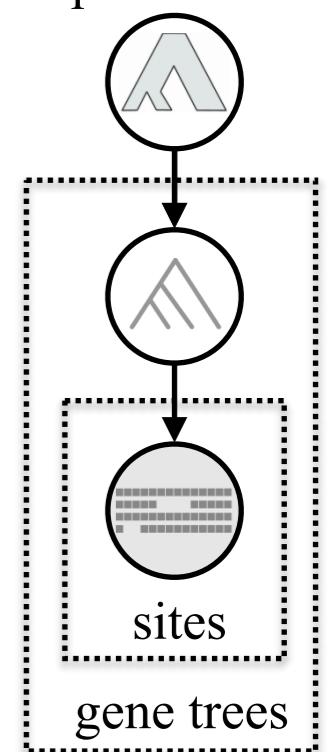
blue.tree



S.tree



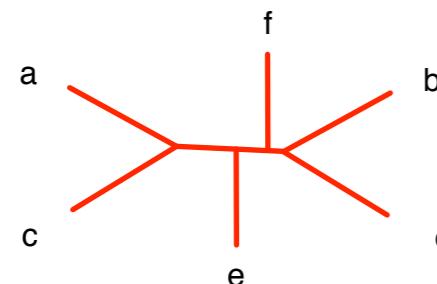
DTL
species tree



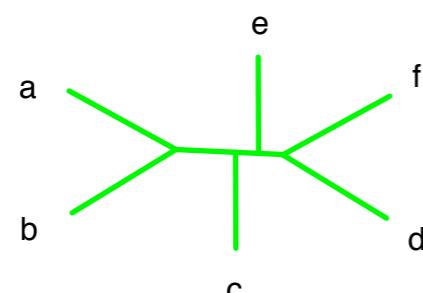
abcdef/

```
.. ./alerax --families red_families.txt --species-tree S.tree --prefix red  
--gene-tree-samples 10 --seed 42 --rec-model UndatedDTL --model-parametrization GLOBAL  
.. ./alerax --families green_families.txt --species-tree S.tree --prefix green  
--gene-tree-samples 10 --seed 42 --rec-model UndatedDTL --model-parametrization GLOBAL  
.. ./alerax --families blue_families.txt --species-tree S.tree --prefix blue  
--gene-tree-samples 10 --seed 42 --rec-model UndatedDTL --model-parametrization GLOBAL
```

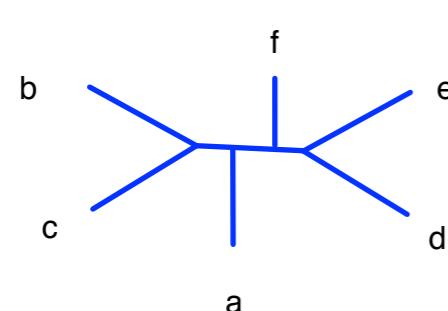
red.tree



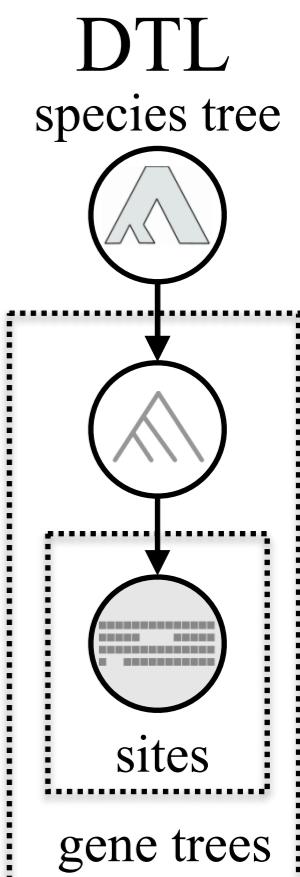
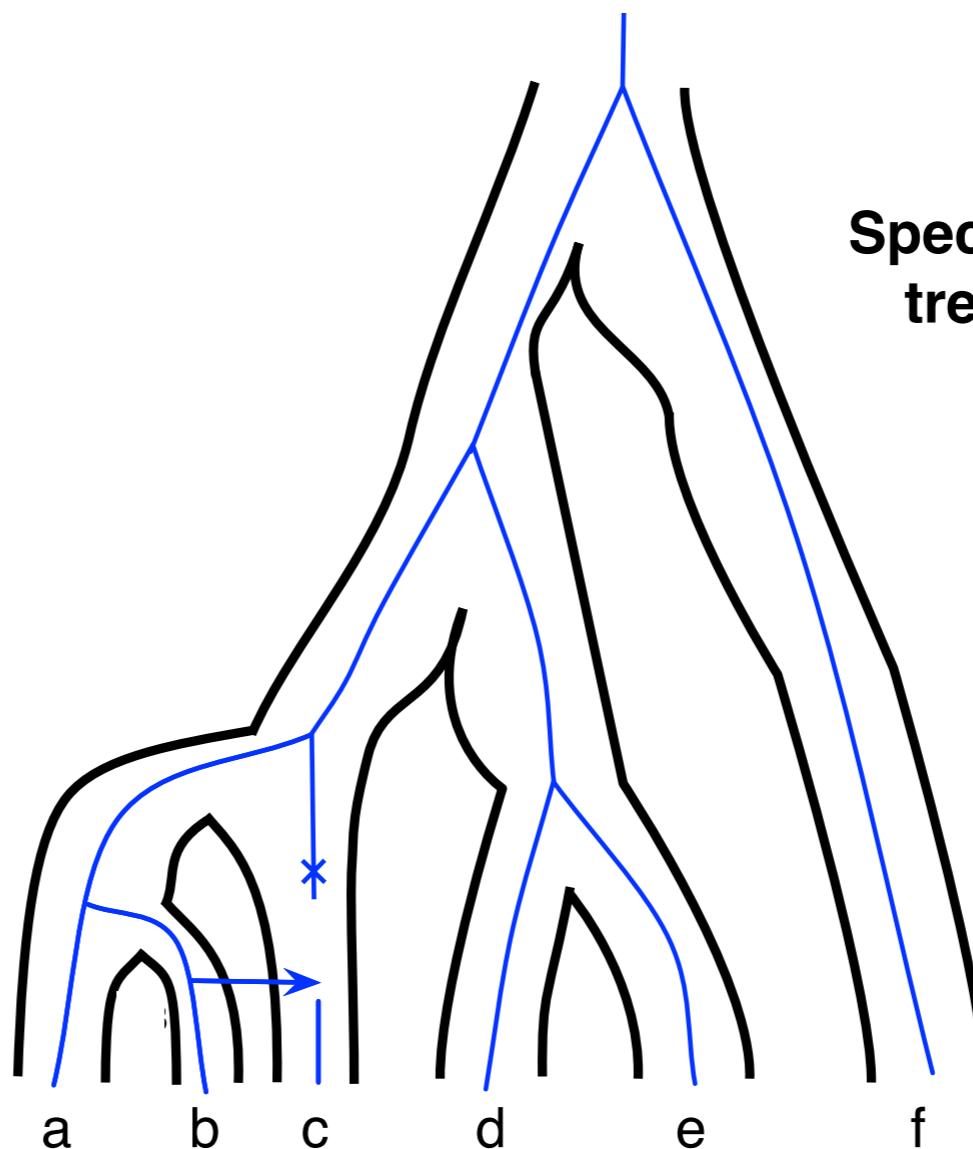
green.tree



blue.tree



Species tree

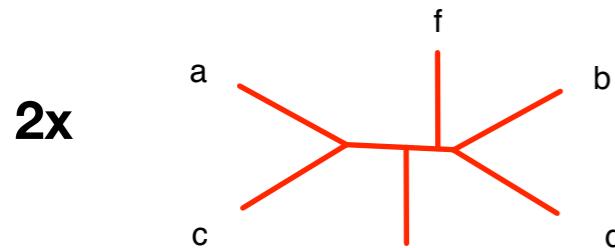


abcdef/

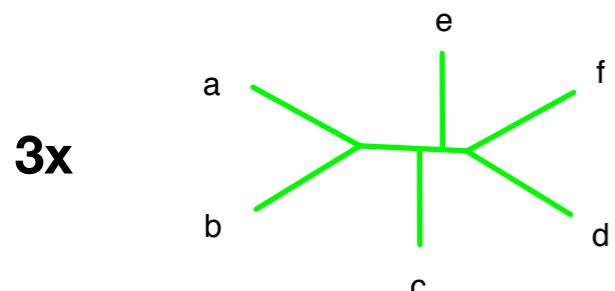
```
./alerax --families rgb_families.txt --species-tree S.tree --prefix rgb  
--gene-tree-samples 10 --seed 42 --rec-model UndatedDTL --model-parametrization GLOBAL
```

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

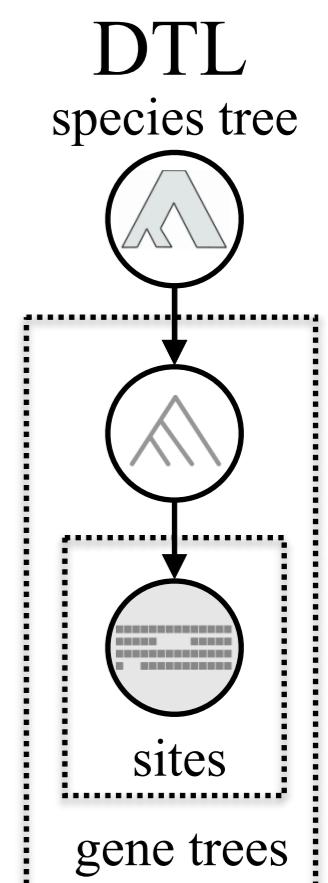
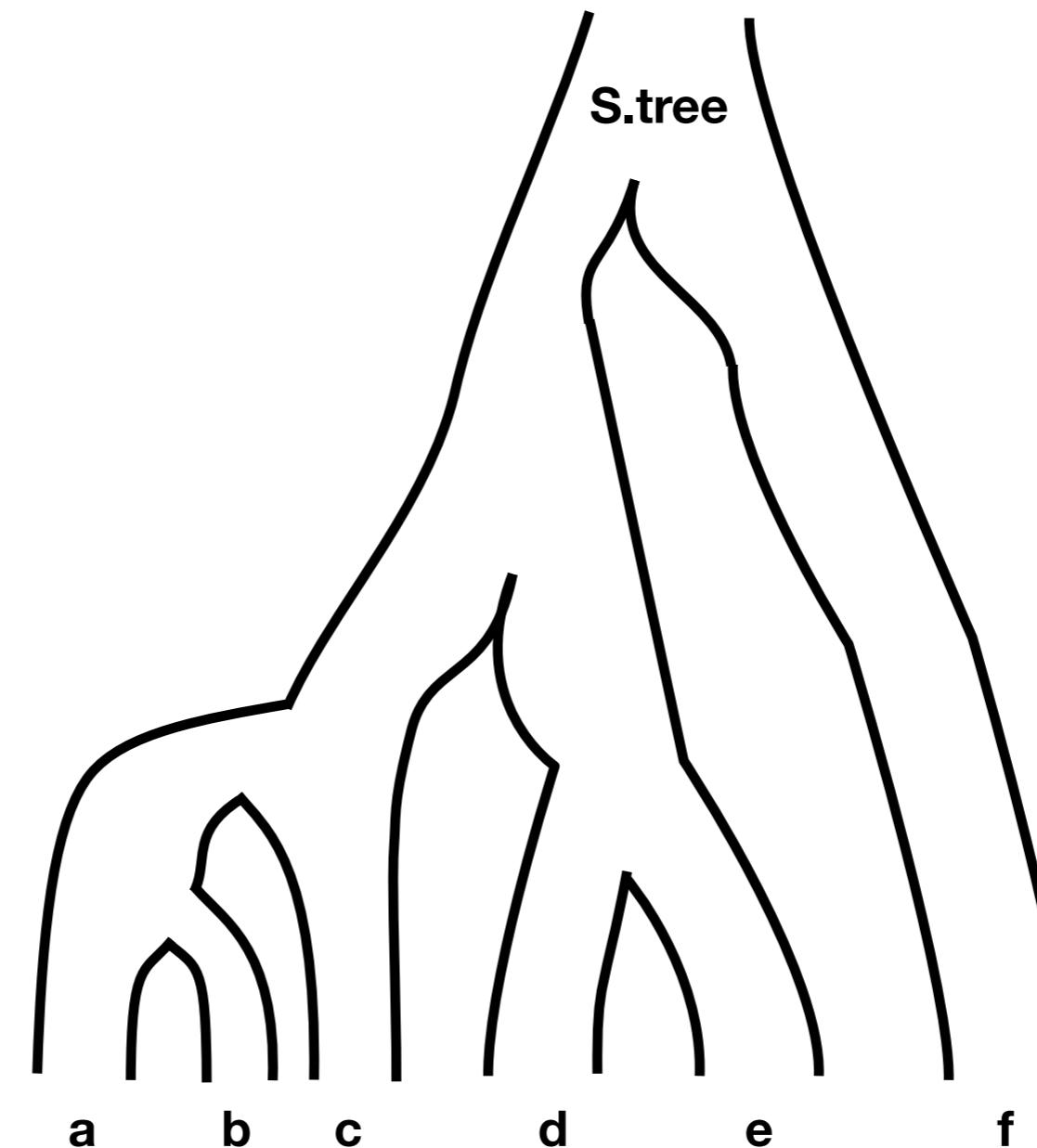
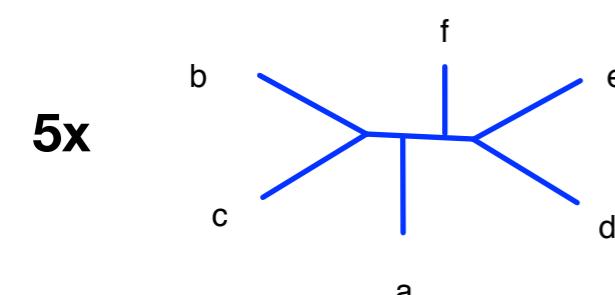
red.tree



green.tree

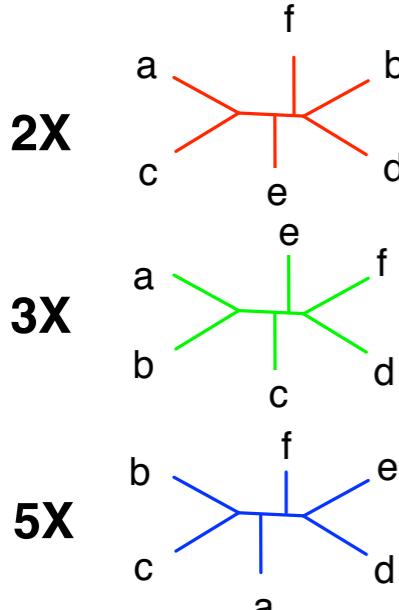


blue.tree

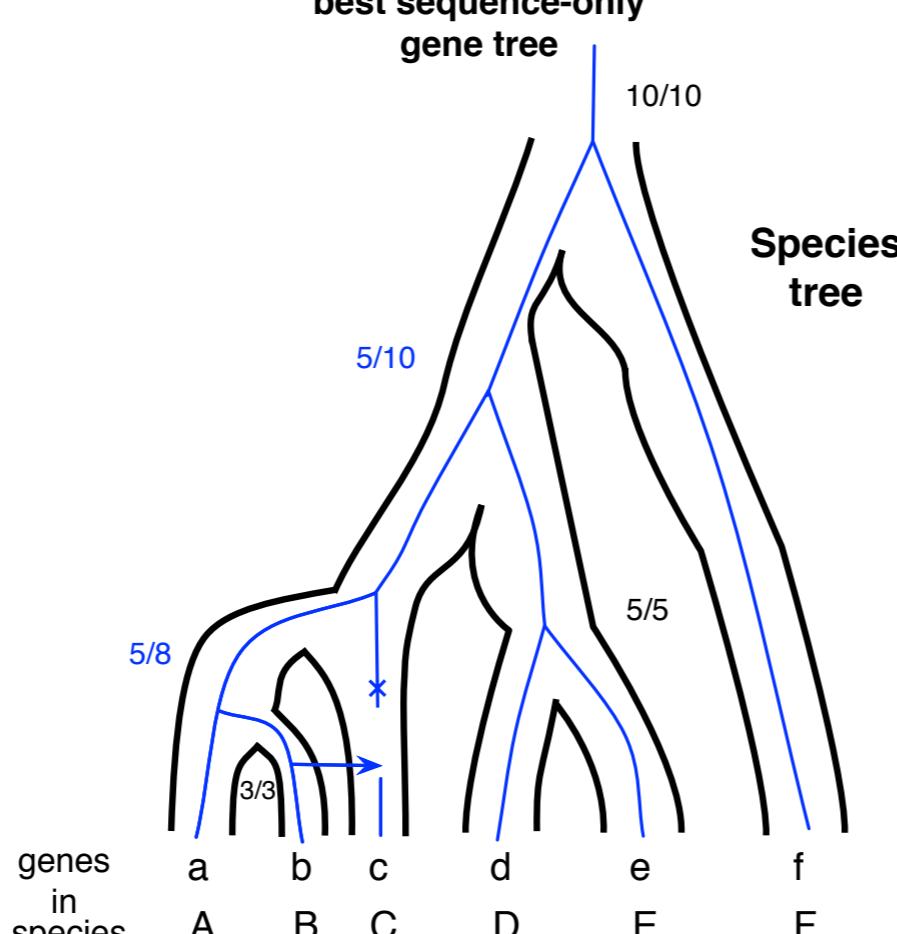


abcdef/

**sample of gene trees
based on sequence**



**best sequence-only
gene tree**

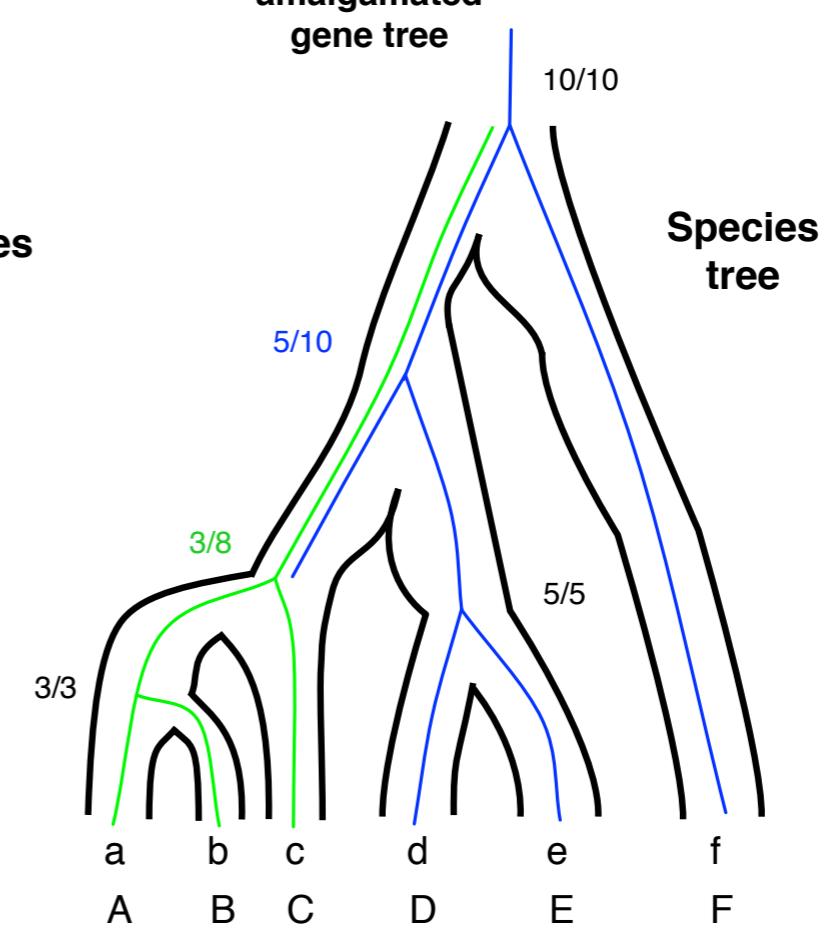


joint likelihoods:

$$\frac{5}{10} \times \frac{5}{8} \times P(1 \text{ transfer and 1 loss})$$

blue -9.49402

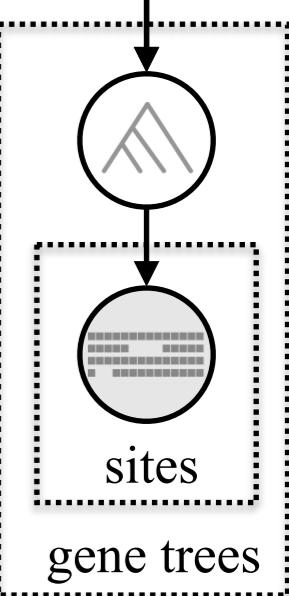
**amalgamated
gene tree**



$$\frac{5}{10} \times \frac{3}{8} \times P(0 \text{ events})$$

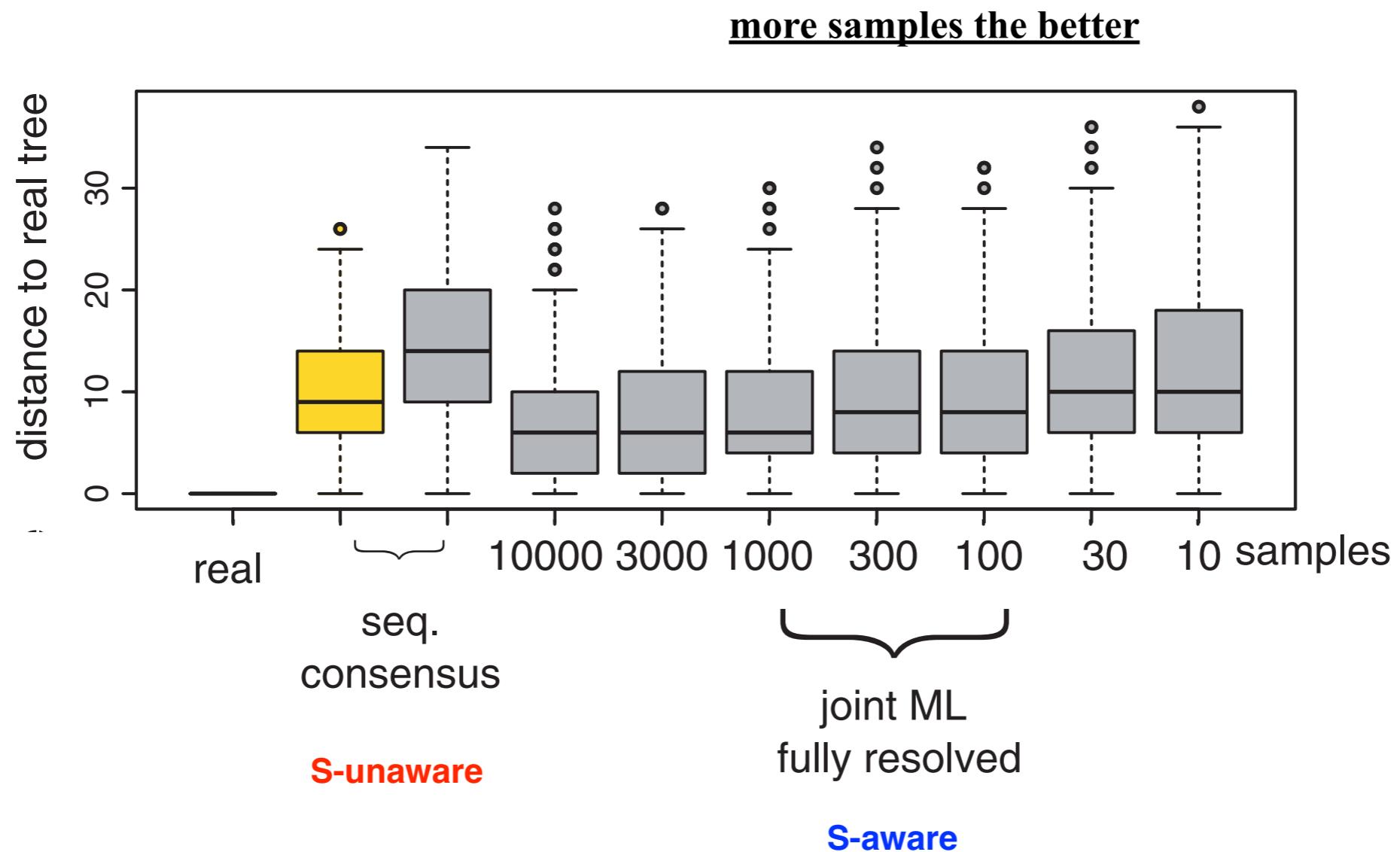
rgb -3.8712

**DTL
species tree**



Efficiently exploring the space of reconciled gene trees

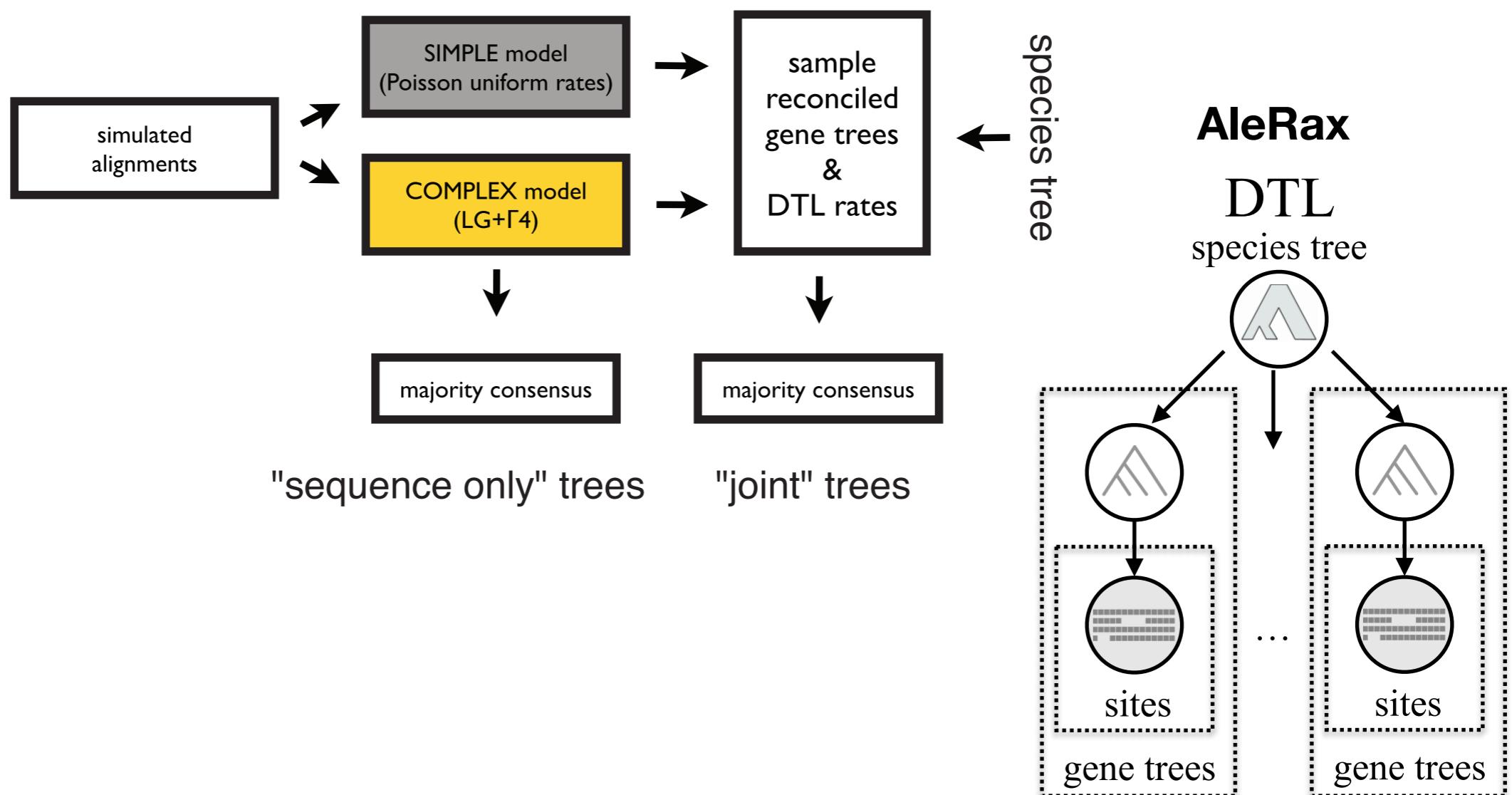
Based on a sample of trees conditional clade probabilities can be used to estimate posterior probability of any gene tree that can be amalgamated. This is usually a very large number of trees (e.g. for 10^4 samples 10^{12} trees, but up to 10^{40}).



real_data/

Real data!

OK .. first realistic data



real_data/

Sample trees using bootstrap and take a look:

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

Run AleRax on the **ML** tree

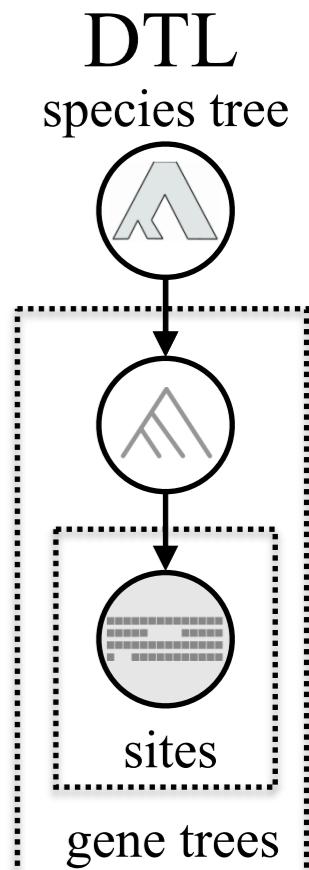
```
../alerax --families HBG486560_ML.families --species-tree S.tree  
--prefix ml --gene-tree-samples 100 --seed 42 --rec-model UndatedDTL  
--model-parametrization GLOBAL
```

Run AleRax on the **true** tree

```
../alerax --families HBG486560_true.families --species-tree S.tree  
--prefix true --gene-tree-samples 100 --seed 42 --rec-model UndatedDTL  
--model-parametrization GLOBAL
```

Run AleRax on the **sample of trees**

```
../alerax --families HBG486560_ufboot.families --species-tree S.tree  
--prefix ufboot --gene-tree-samples 100 --seed 42 --rec-model UndatedDTL  
--model-parametrization GLOBAL
```



real_data/

Sample trees using bootstrap and take a look:

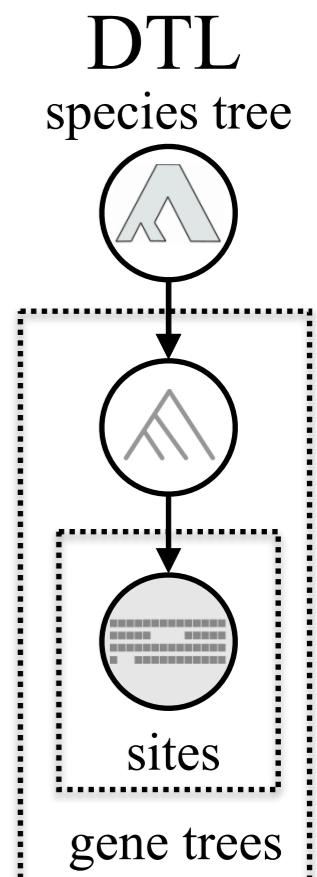
```
iqtree2 -m LG -bb 10000 -s HBG486560_real.fasta -wbtl  
less HBG486560_real.fasta.ufboot
```

Run AleRax on the ML tree

```
../alerax --families HBG486560_real_ML.families --species-tree S.tree  
--prefix real_ml --gene-tree-samples 100 --seed 42 --rec-model UndatedDTL  
--model-parametrization GLOBAL
```

Run AleRax on the **sample of trees**

```
../alerax --families HBG486560_real_ufboot.families --species-tree S.tree  
--prefix real_ufboot --gene-tree-samples 100 --seed 42 --rec-model UndatedDTL  
--model-parametrization GLOBAL
```



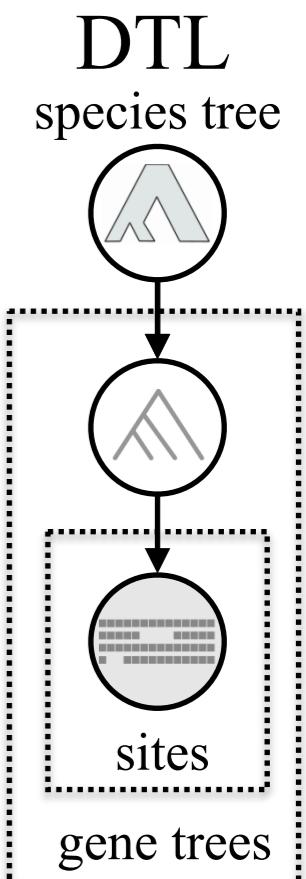
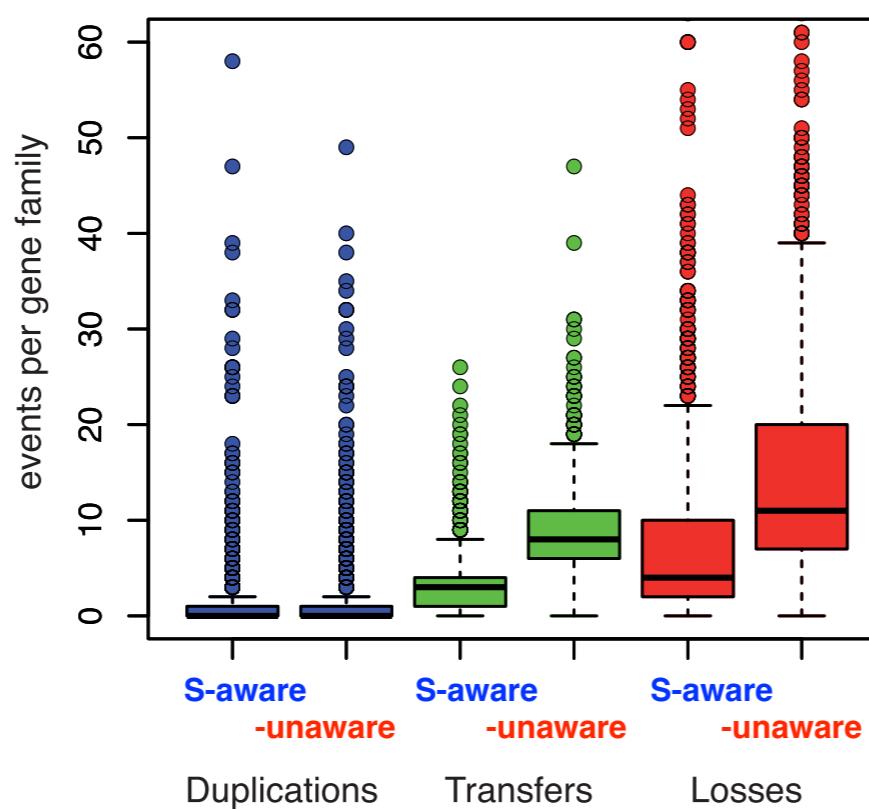
real_data/

./alerax --families HBG486560_real_ML.families [..]

Spec	Dupl	Loss	Tra	Pres	Orig	Cop	Single
68.8	0	3.85	6.05	68.57	1	68.8	65.56

./alerax --families HBG486560_real_ufboot.families [..]

Spec	Dupl	Loss	Tra	Pres	Orig	Cop	Single
71.01	0	1.02	1.01	70.99	1	71.01	71



Can you find a true orthologous family?

(i.e. a family with 0 Duplication , Transfer and Loss events)