An overview of relaxed molecular clock models

Marc Manceau, Amaury Lambert, Hélène Morlon

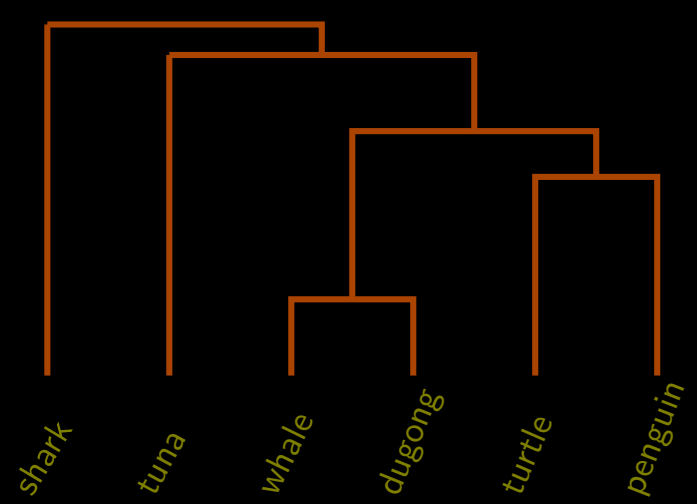
December 1, 2017

The first intuition

Back in the 60', with Zuckerkandl and Pauling

- Pairwise differences are compatible with branch lengths of an ultrametric tree.
- This is a fictious example. Historical data probably looked more like 'chicken, mouse, drosophilia, C. elegans'...

	shark	tuna	whale	dugong	turtle	penguin
shark	0	46	47	45	48	44
tuna		0	43	42	39	41
whale			0	9	30	33
dugong				0	32	34
turtle					0	27
penguin						0



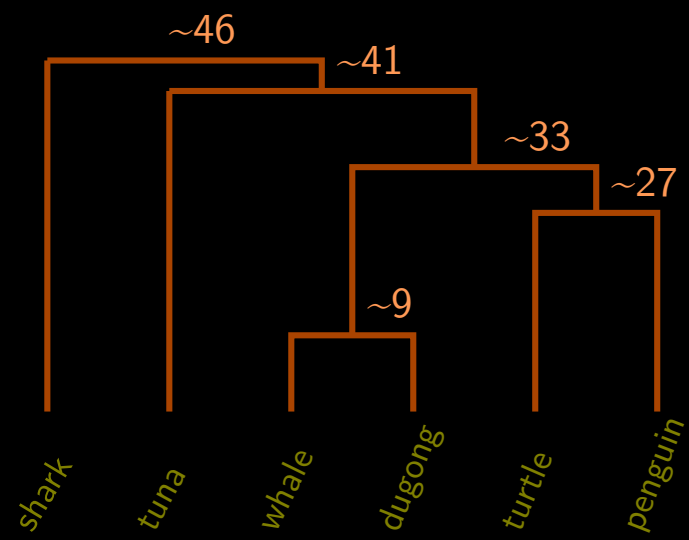
- Mutations modeled as a Poisson process with constant rate on the tree.

The first intuition

Back in the 60', with Zuckerkandl and Pauling

- Pairwise differences are compatible with branch lengths of an ultrametric tree.
- This is a fictious example. Historical data probably looked more like 'chicken, mouse, drosophilia, C. elegans'...

	shark	tuna	whale	dugong	turtle	penguin
shark	0	46	47	45	48	44
tuna		0	43	42	39	41
whale			0	9	30	33
dugong				0	32	34
turtle					0	27
penguin						0

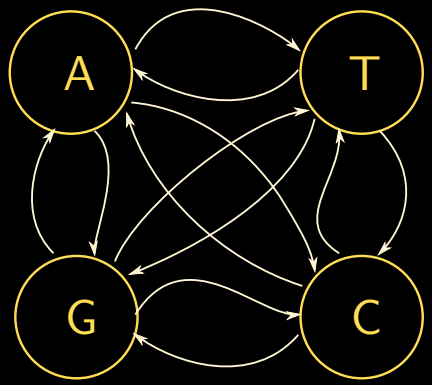


- Mutations modeled as a Poisson process with constant rate on the tree.

The first intuition

Simulation of molecular evolution

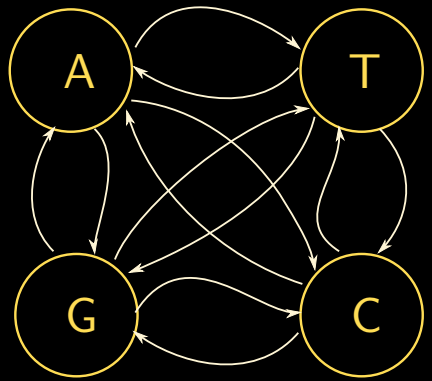
- The model comes with a transition matrix $P(t)$ and a stationary distribution π.
- All nucleotides at the root are iid $\sim \pi$.
- The final state of each nucleotide on each branch is drawn using $P(t)$.
- The process is copied independently on sister branches.

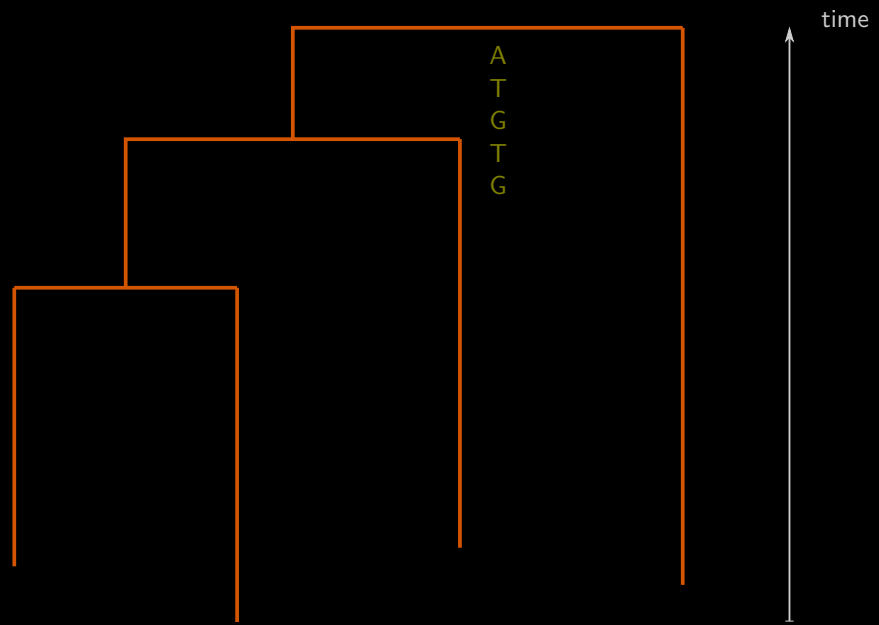


The first intuition

Simulation of molecular evolution

- The model comes with a transition matrix $P(t)$ and a stationary distribution π.
- All nucleotides at the root are iid $\sim \pi$.
- The final state of each nucleotide on each branch is drawn using $P(t)$.
- The process is copied independently on sister branches.

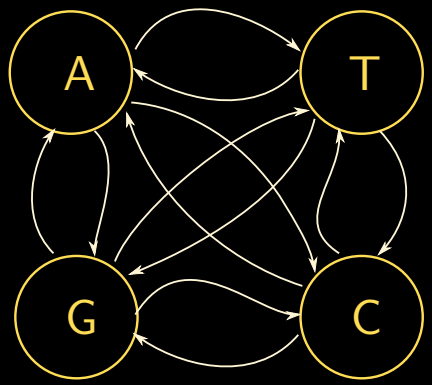


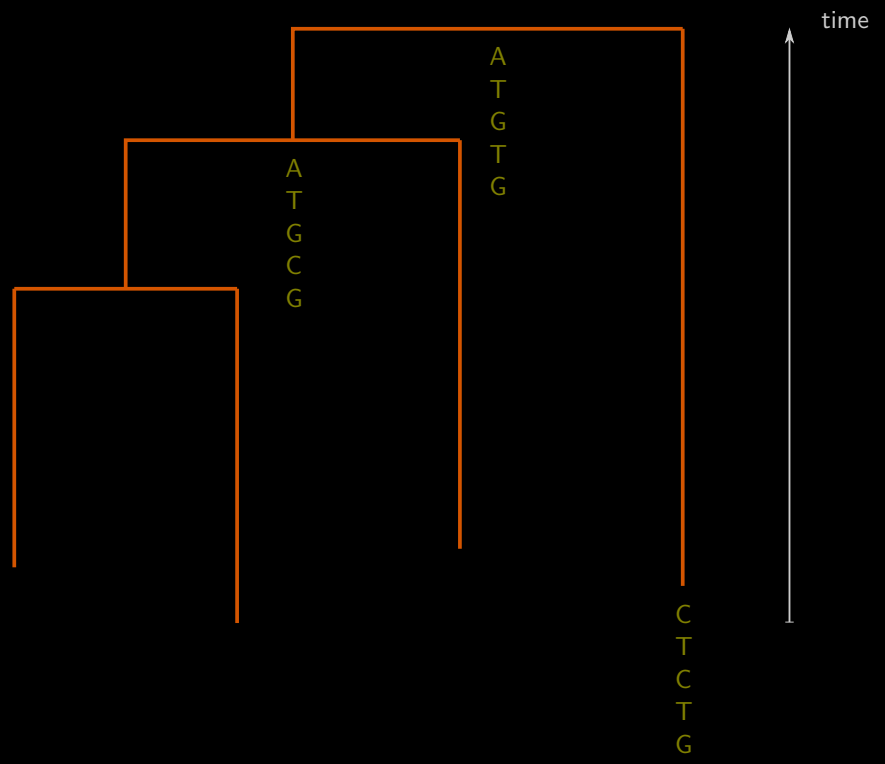


The first intuition

Simulation of molecular evolution

- The model comes with a transition matrix $P(t)$ and a stationary distribution π.
- All nucleotides at the root are iid $\sim \pi$.
- The final state of each nucleotide on each branch is drawn using $P(t)$.
- The process is copied independently on sister branches.

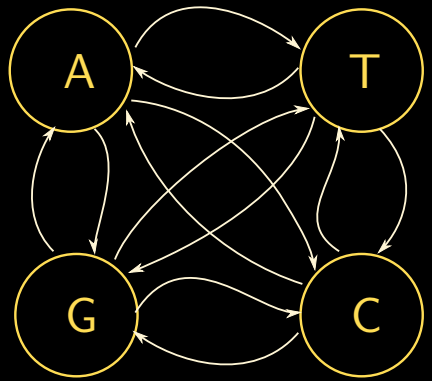


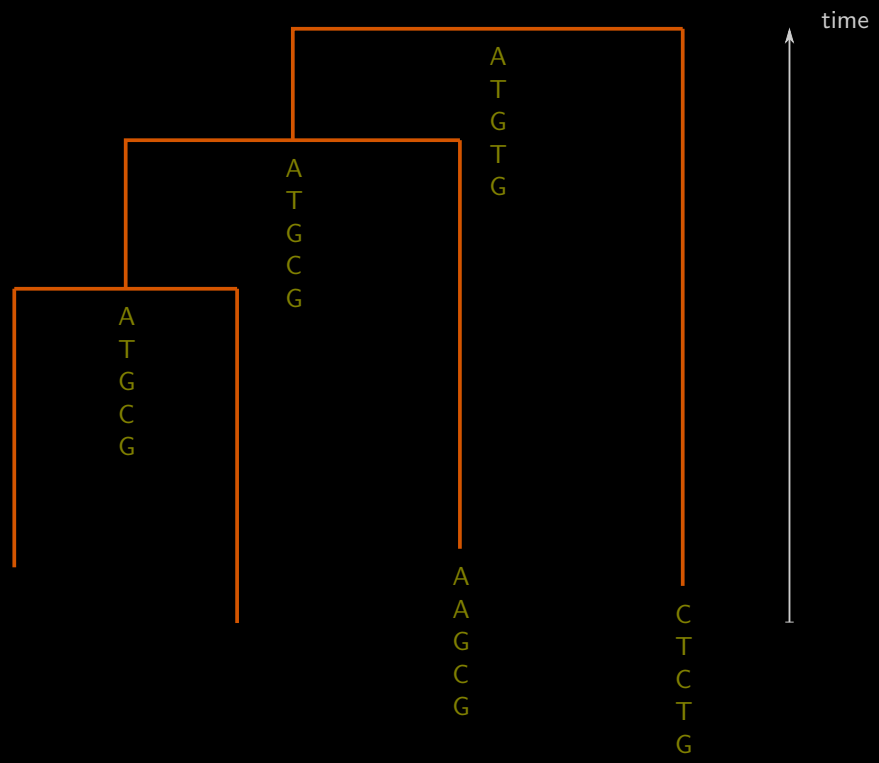


The first intuition

Simulation of molecular evolution

- The model comes with a transition matrix $P(t)$ and a stationary distribution π.
- All nucleotides at the root are iid $\sim \pi$.
- The final state of each nucleotide on each branch is drawn using $P(t)$.
- The process is copied independently on sister branches.

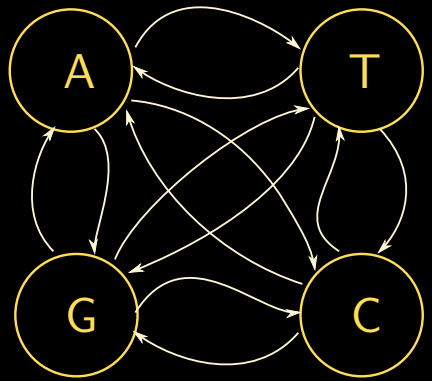


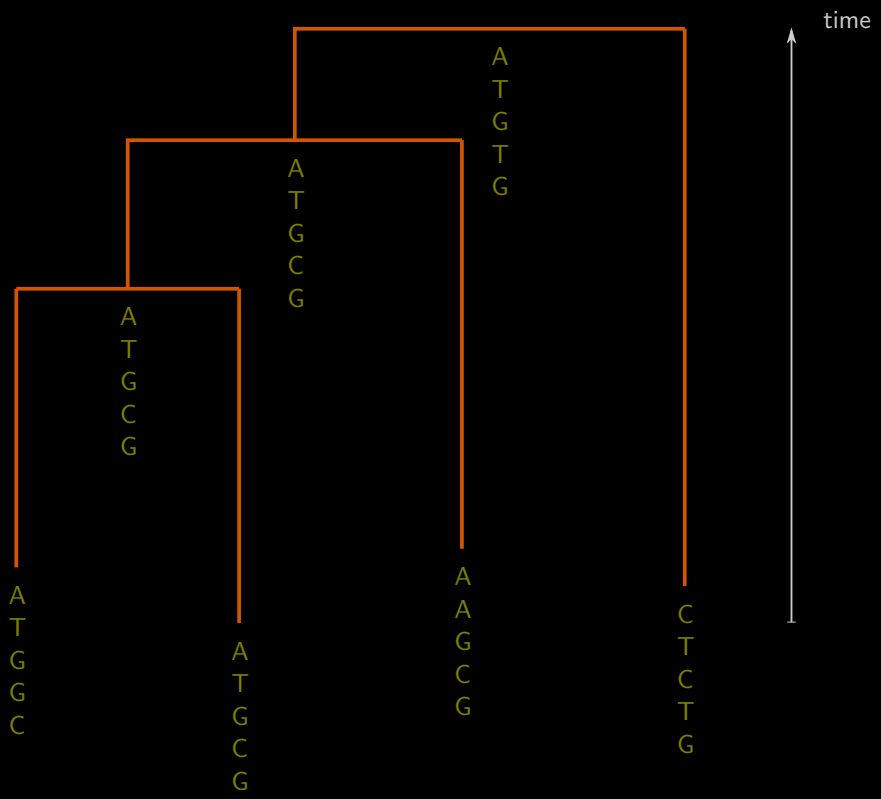


The first intuition

Simulation of molecular evolution

- The model comes with a transition matrix $P(t)$ and a stationary distribution π.
- All nucleotides at the root are iid $\sim \pi$.
- The final state of each nucleotide on each branch is drawn using $P(t)$.
- The process is copied independently on sister branches.

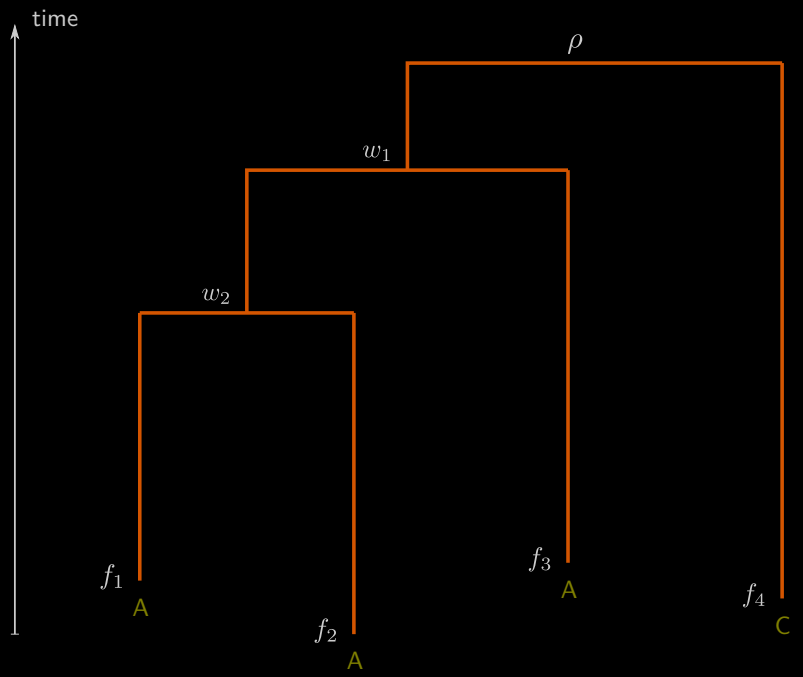




The inference framework

Likelihood computation (Felsenstein in the 80')

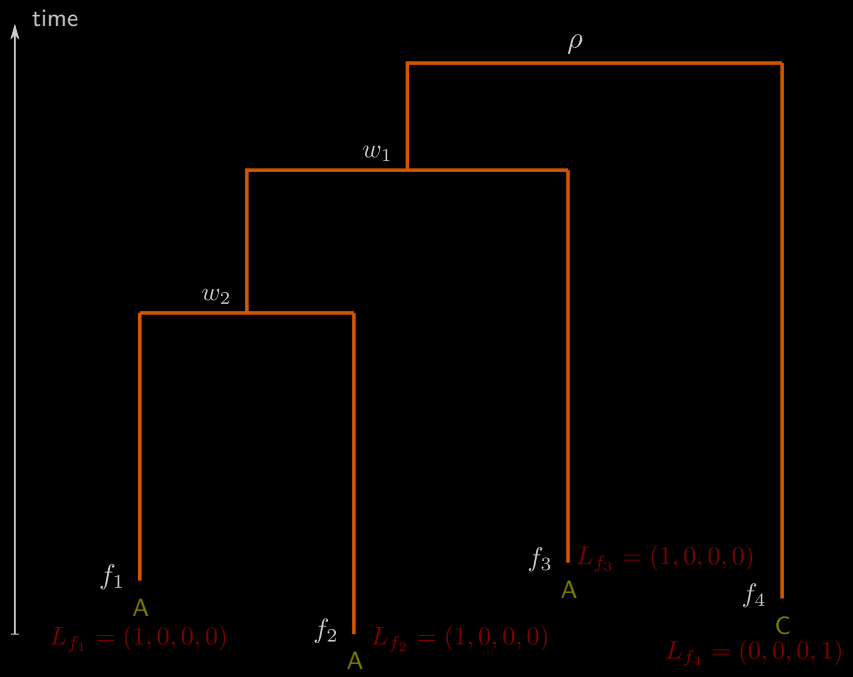
- We call X_w, the nucleotide state at node w.
  We further define $L_w := (\mathbb{P}(\text{ tip data } \mid X_w = i))_{i \in \{A,T,G,C\}}$
- On a leaf f, initialize $L_f = (\mathbb{1}_{X_f=A}, \mathbb{1}_{X_f=T}, \mathbb{1}_{X_f=G}, \mathbb{1}_{X_f=C})$
- On a node (e.g. w_2) having descent f_1 and f_2, $L_{w_2} = (P(t_1)L_{f_1}) \cdot (P(t_2)L_{f_2})$
  (where $\cdot$ is the Hadamard product)
- At the root, $L = \pi L_\rho$.



The inference framework

Likelihood computation (Felsenstein in the 80')

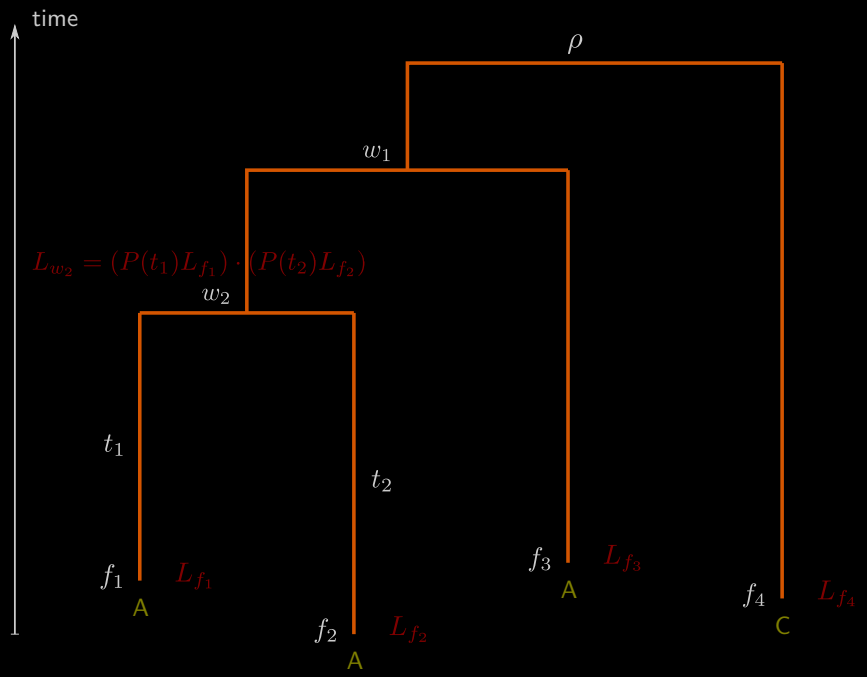
- We call X_w, the nucleotide state at node w.
  We further define $L_w := (\mathbb{P}(\text{ tip data } \mid X_w = i))_{i \in \{A,T,G,C\}}$
- On a leaf f, initialize $L_f = (\mathbb{1}_{X_f=A}, \mathbb{1}_{X_f=T}, \mathbb{1}_{X_f=G}, \mathbb{1}_{X_f=C})$
- On a node (e.g. w_2) having descent f_1 and f_2, $L_{w_2} = (P(t_1)L_{f_1}) \cdot (P(t_2)L_{f_2})$ (where $\cdot$ is the Hadamard product)
- At the root, $L = \pi L_\rho$.



The inference framework

Likelihood computation (Felsenstein in the 80')

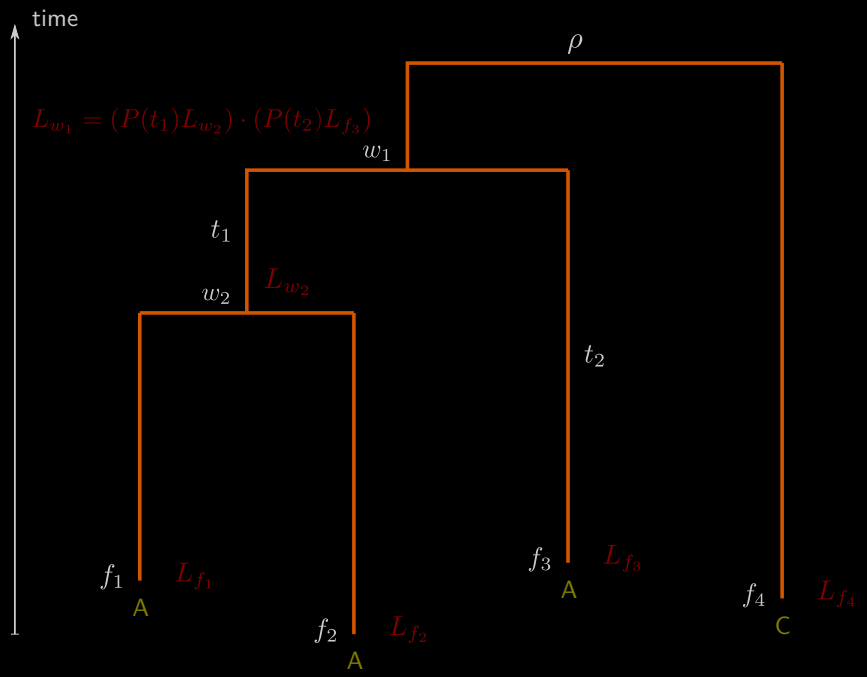
- We call X_w, the nucleotide state at node w.
  We further define $L_w := (\mathbb{P}(\text{ tip data } \mid X_w = i))_{i \in \{A,T,G,C\}}$
- On a leaf f, initialize $L_f = (\mathbb{1}_{X_f=A}, \mathbb{1}_{X_f=T}, \mathbb{1}_{X_f=G}, \mathbb{1}_{X_f=C})$
- On a node (e.g. w_2) having descent f_1 and f_2, $L_{w_2} = (P(t_1)L_{f_1}) \cdot (P(t_2)L_{f_2})$
  (where $\cdot$ is the Hadamard product)
- At the root, $L = \pi L_\rho$.



The inference framework

Likelihood computation (Felsenstein in the 80')

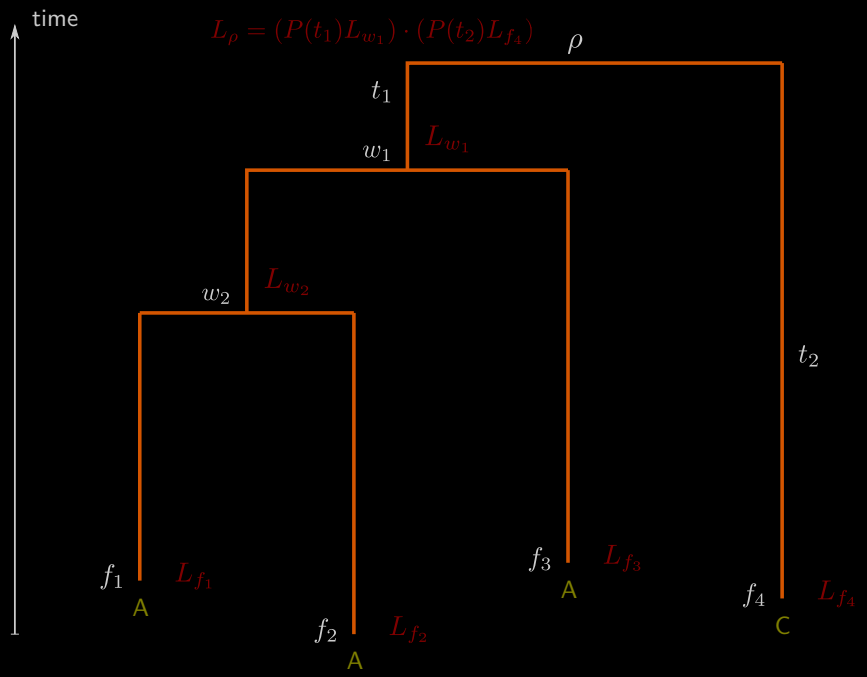
- We call X_w, the nucleotide state at node w.
  We further define $L_w := (\mathbb{P}(\text{ tip data } \mid X_w = i))_{i \in \{A,T,G,C\}}$
- On a leaf f, initialize $L_f = (\mathbb{1}_{X_f=A}, \mathbb{1}_{X_f=T}, \mathbb{1}_{X_f=G}, \mathbb{1}_{X_f=C})$
- On a node (e.g. w_2) having descent f_1 and f_2, $L_{w_2} = (P(t_1)L_{f_1}) \cdot (P(t_2)L_{f_2})$
  (where $\cdot$ is the Hadamard product)
- At the root, $L = \pi L_\rho$.



The inference framework

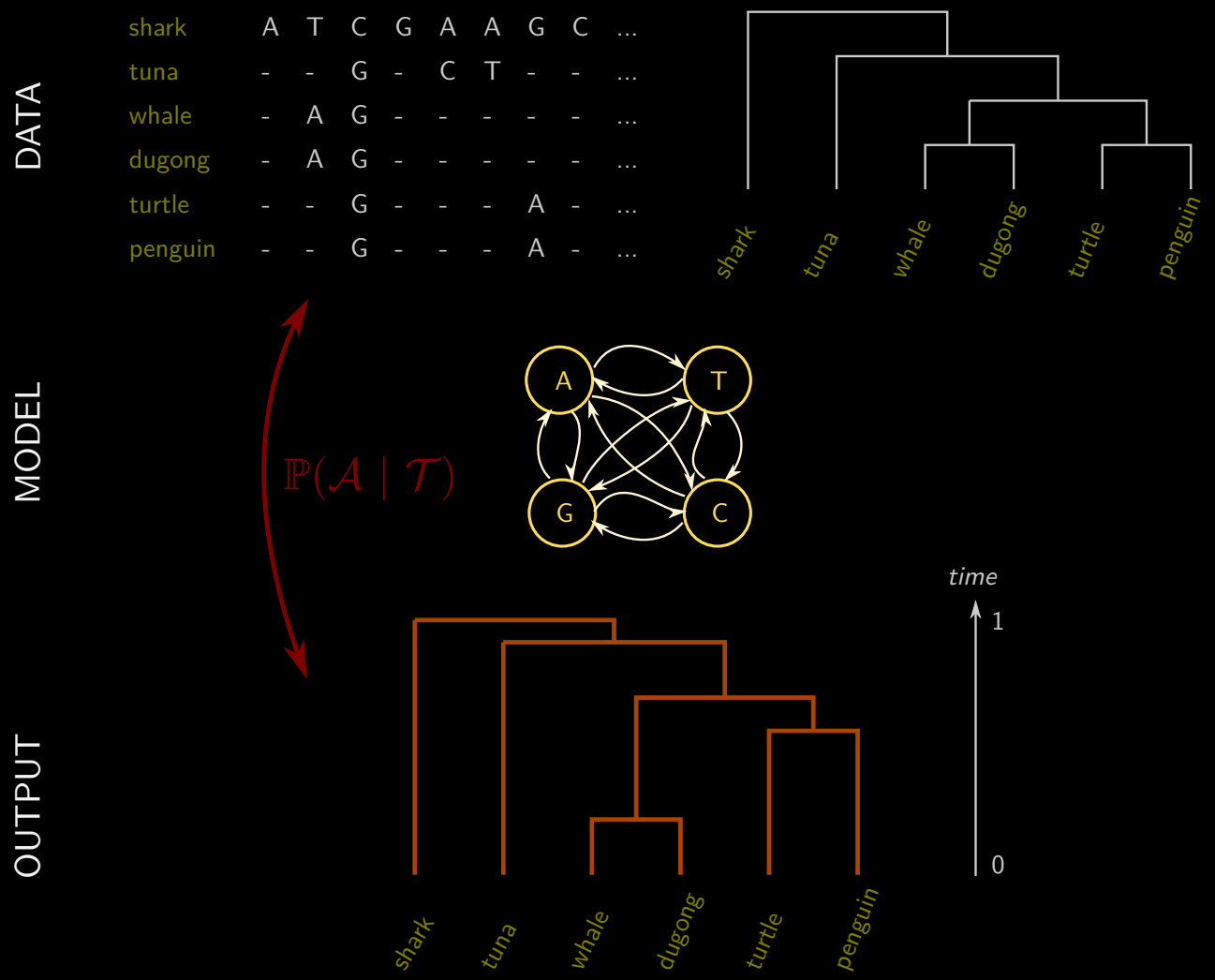
Likelihood computation (Felsenstein in the 80')

- We call X_w, the nucleotide state at node w.
  We further define $L_w := (\mathbb{P}(\text{ tip data } \mid X_w = i))_{i \in \{A,T,G,C\}}$
- On a leaf f, initialize $L_f = (\mathbb{1}_{X_f=A}, \mathbb{1}_{X_f=T}, \mathbb{1}_{X_f=G}, \mathbb{1}_{X_f=C})$
- On a node (e.g. w_2) having descent f_1 and f_2, $L_{w_2} = (P(t_1)L_{f_1}) \cdot (P(t_2)L_{f_2})$
  (where $\cdot$ is the Hadamard product)
- At the root, $L = \pi L_\rho$.



The inference framework

Need for models of molecular evolution

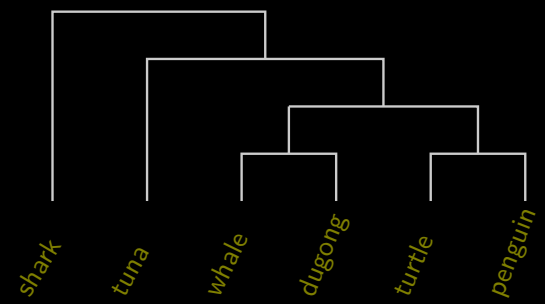


The inference framework

Need for models of molecular evolution

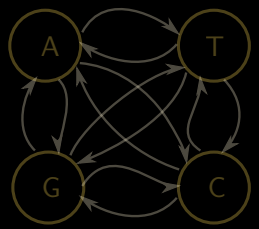
DATA

shark	A	T	C	G	A	A	G	C	...
tuna	-	-	G	-	C	T	-	-	...
whale	-	A	G	-	-	-	-	-	...
dugong	-	A	G	-	-	-	-	-	...
turtle	-	-	G	-	-	-	A	-	...
penguin	-	-	G	-	-	-	A	-	...



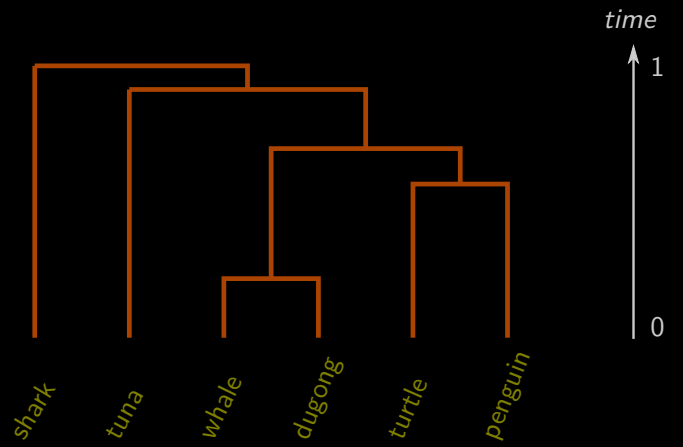
MODEL

$\mathbb{P}(\mathcal{A} \mid \mathcal{T})$



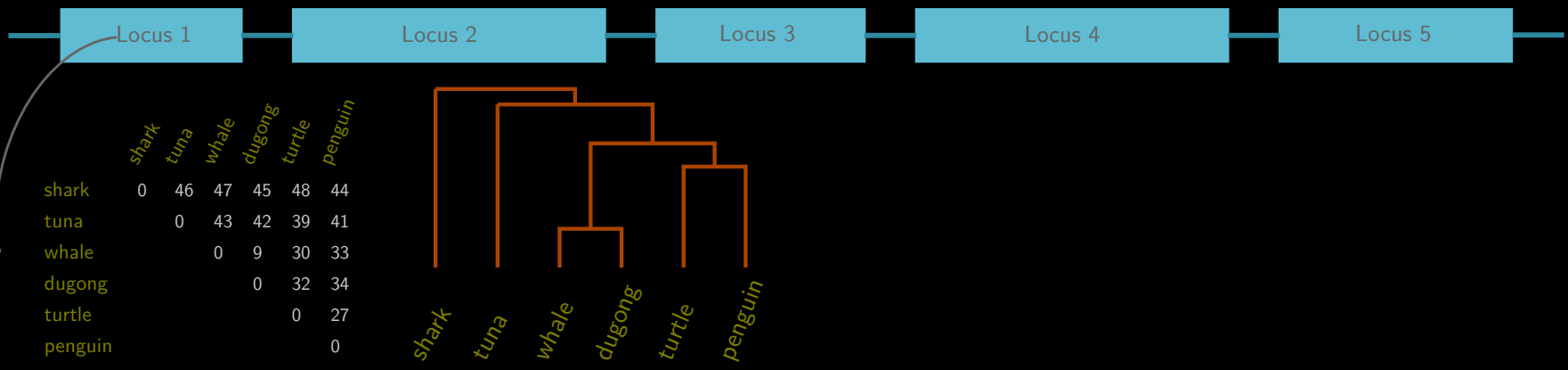
Jukes-Cantor,
Kimura,
Tamura-Nei,
...
Generalized Time-Reversible Model.

OUTPUT



Progress for our understanding of evolution

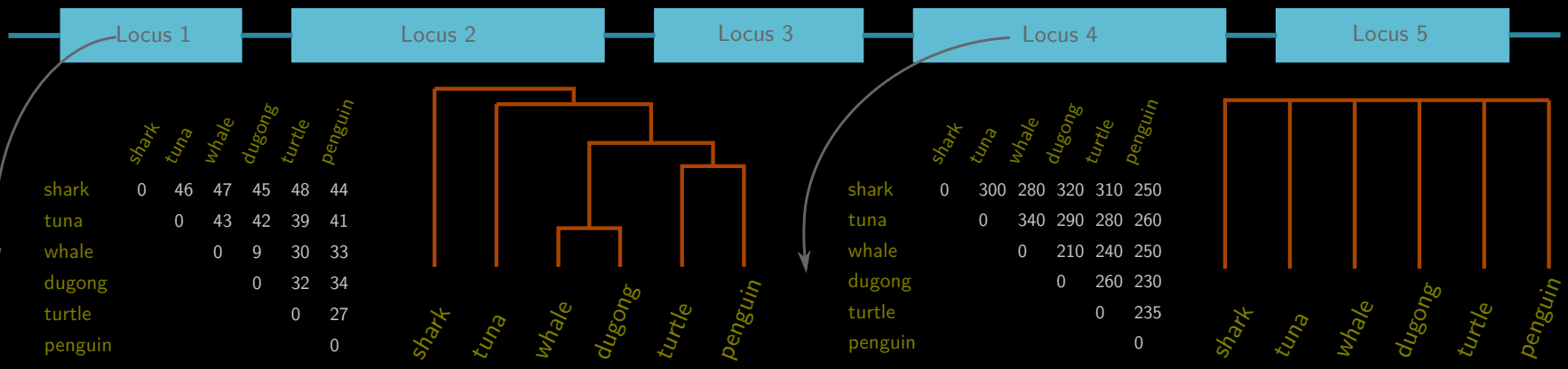
Slow versus fast loci



- Studies on substitution rates across loci.
- Discussion on stabilizing selection and neutrality.
- Widespread tool for relative dating of phylogenies.

Progress for our understanding of evolution

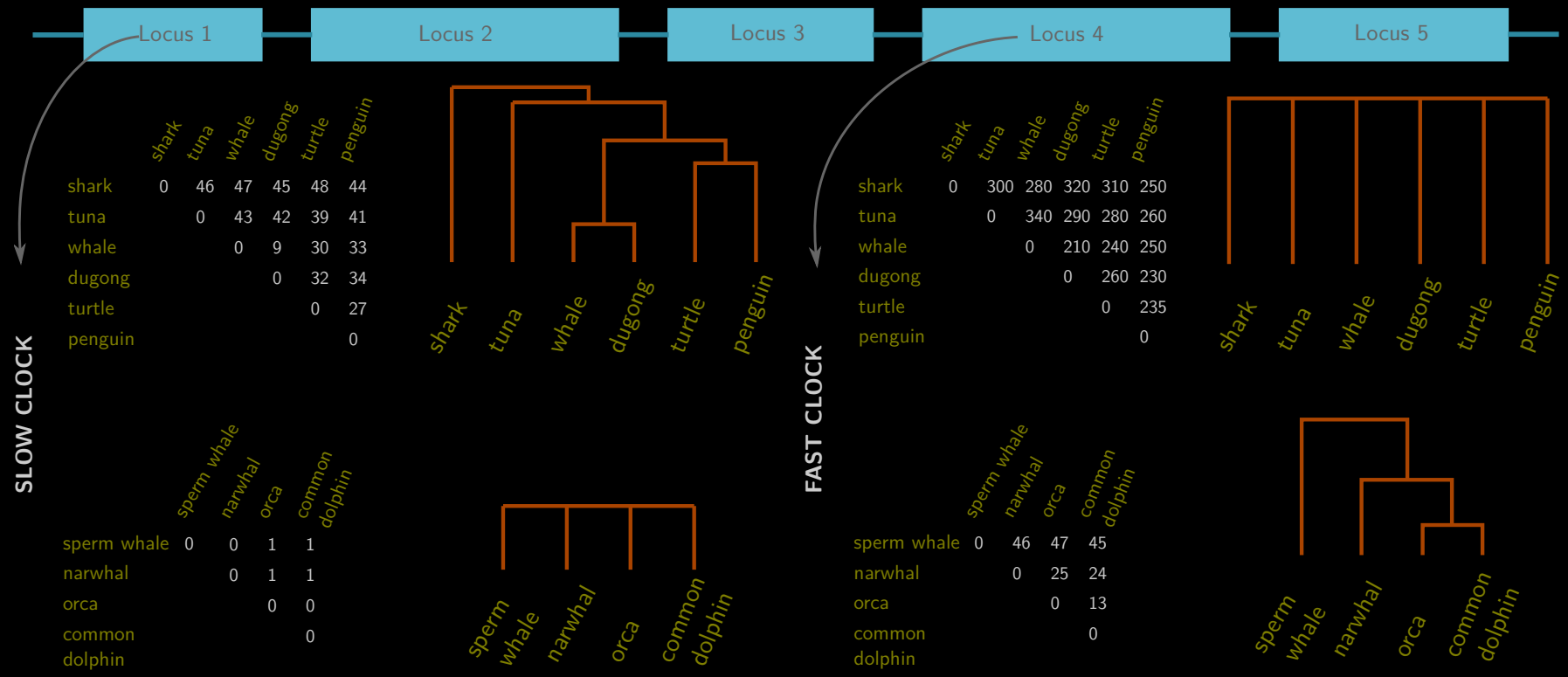
Slow versus fast loci



- Studies on substitution rates across loci.
- Discussion on stabilizing selection and neutrality.
- Widespread tool for relative dating of phylogenies.

Progress for our understanding of evolution

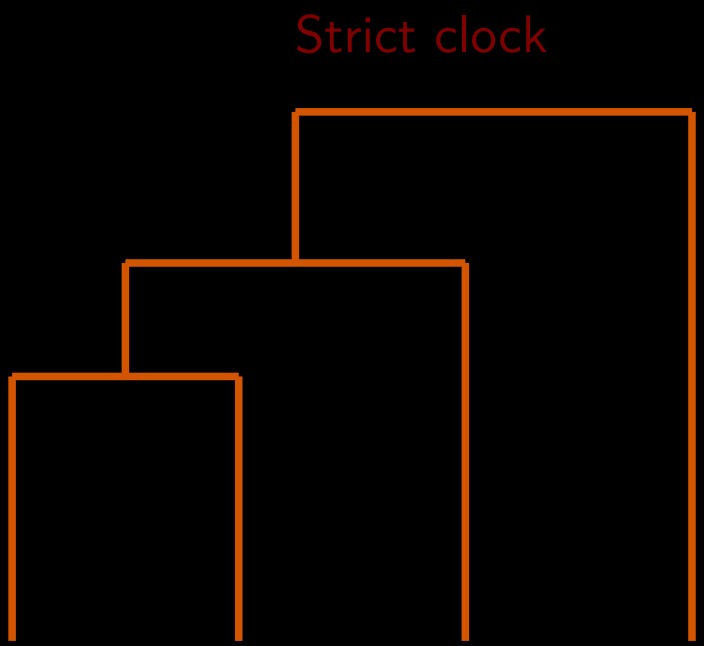
Slow versus fast loci



- Studies on substitution rates across loci.
- Discussion on stabilizing selection and neutrality.
- Widespread tool for relative dating of phylogenies.

Progress for our understanding of evolution

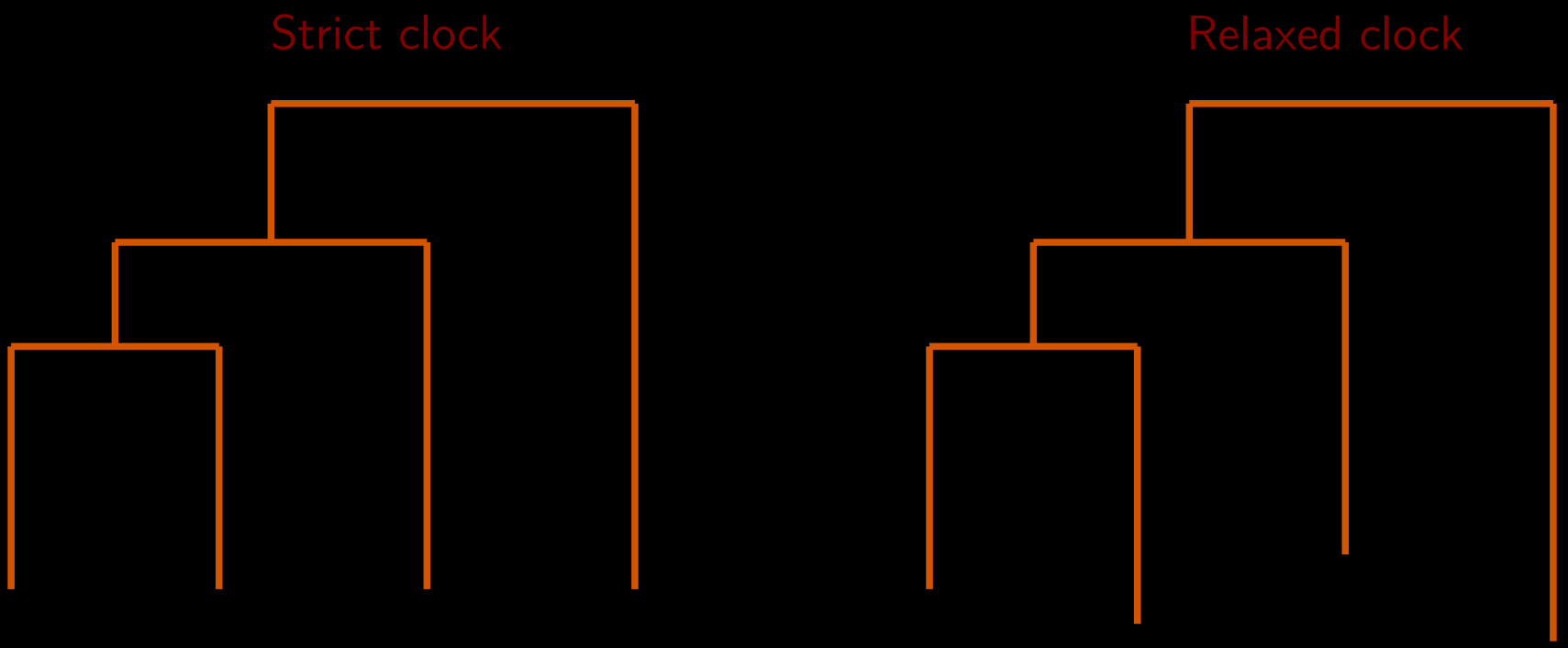
The molecular clock vs. fully relaxed clock



- Either all coalescence times are free parameters ($n-1$).
- Or all branch-lengths are parameters ($2n-3$).
- The relative fit allows for the first tests of the molecular clock hypothesis.

Progress for our understanding of evolution

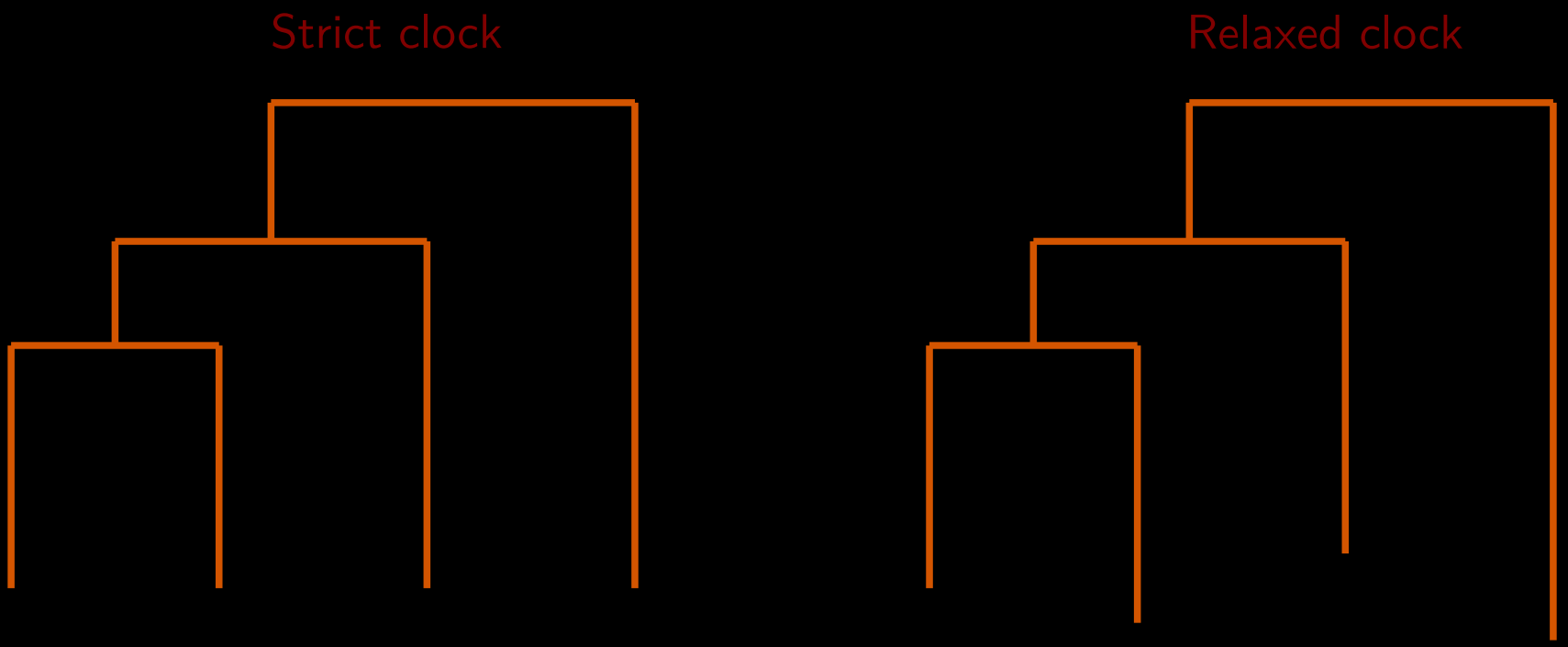
The molecular clock vs. fully relaxed clock



- Either all coalescence times are free parameters ($n-1$).
- Or all branch-lengths are parameters ($2n-3$).
- The relative fit allows for the first tests of the molecular clock hypothesis.

Progress for our understanding of evolution

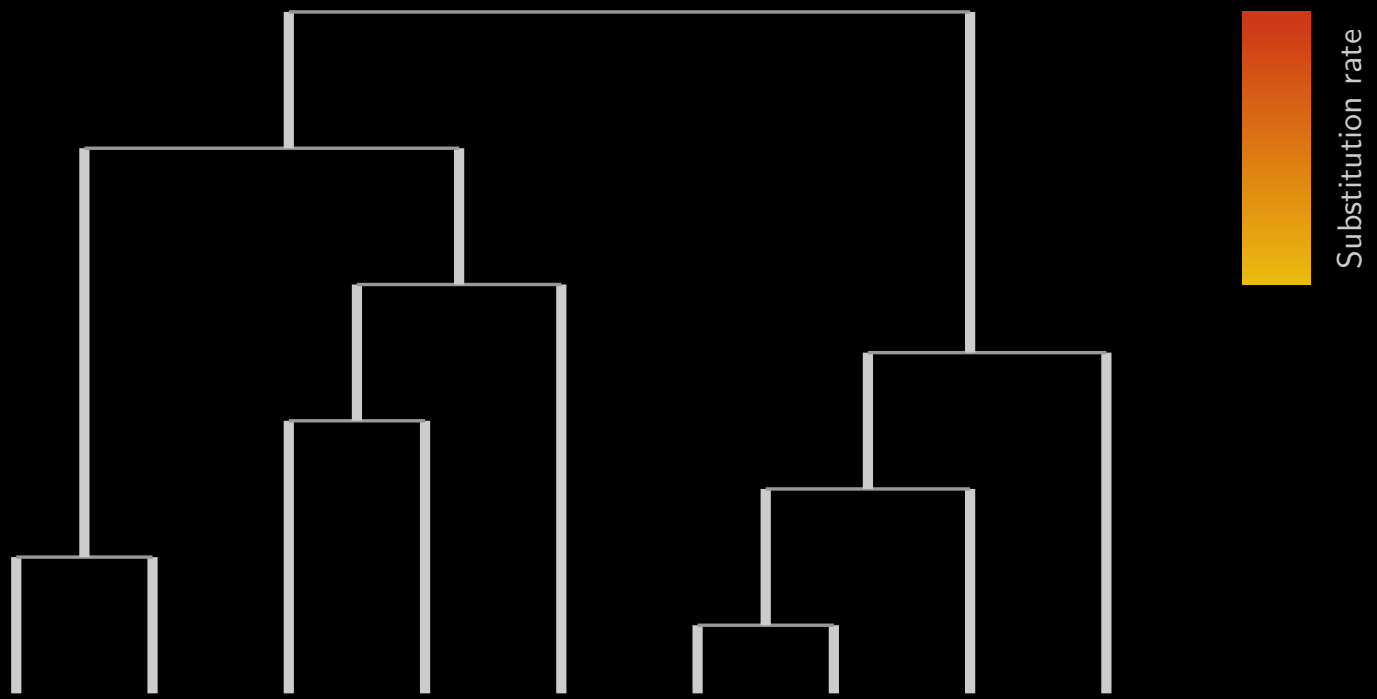
The molecular clock vs. fully relaxed clock



- Either all coalescence times are free parameters ($n-1$).
- Or all branch-lengths are parameters ($2n-3$).
- The relative fit allows for the first tests of the molecular clock hypothesis.

Non-auto-correlated v.s. auto-correlated relaxed clocks

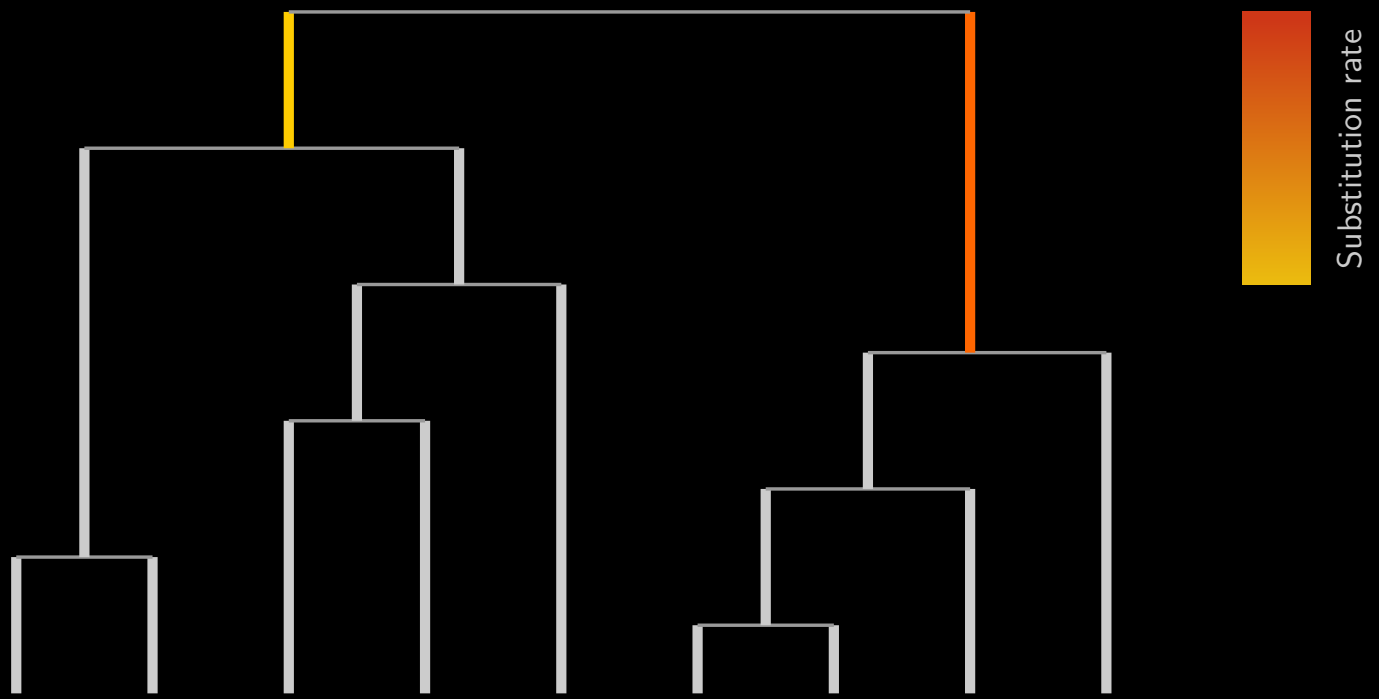
The non-auto-correlated ones first



- The substitution rate is different on each branch.
- Or the changing points are drawn first on the tree (Poisson process).
- *Rate values are chosen in a fixed law, independently of the neighbouring ones.*
- *Rate values depend on the rate in the parent branch.*

Non-auto-correlated v.s. auto-correlated relaxed clocks

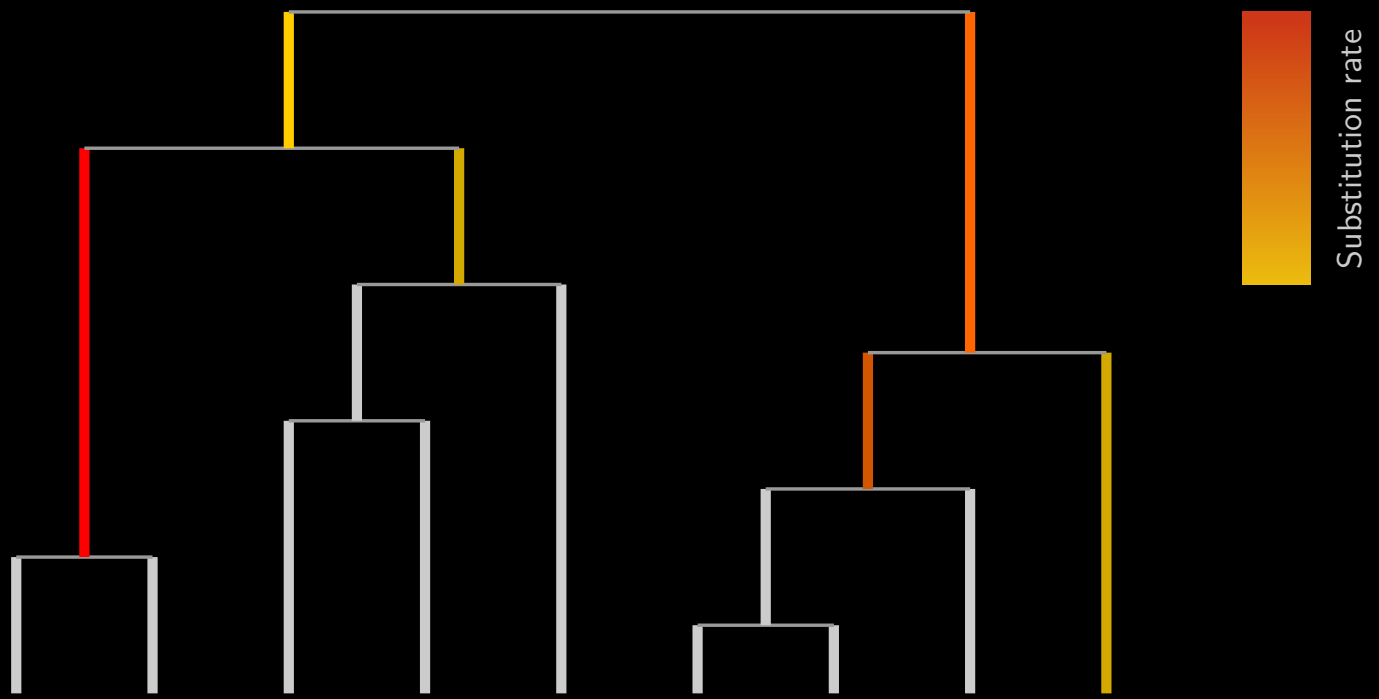
The non-auto-correlated ones first



- The substitution rate is different on each branch.
- Or the changing points are drawn first on the tree (Poisson process).
- *Rate values are chosen in a fixed law, independently of the neighbouring ones.*
- *Rate values depend on the rate in the parent branch.*

Non-auto-correlated v.s. auto-correlated relaxed clocks

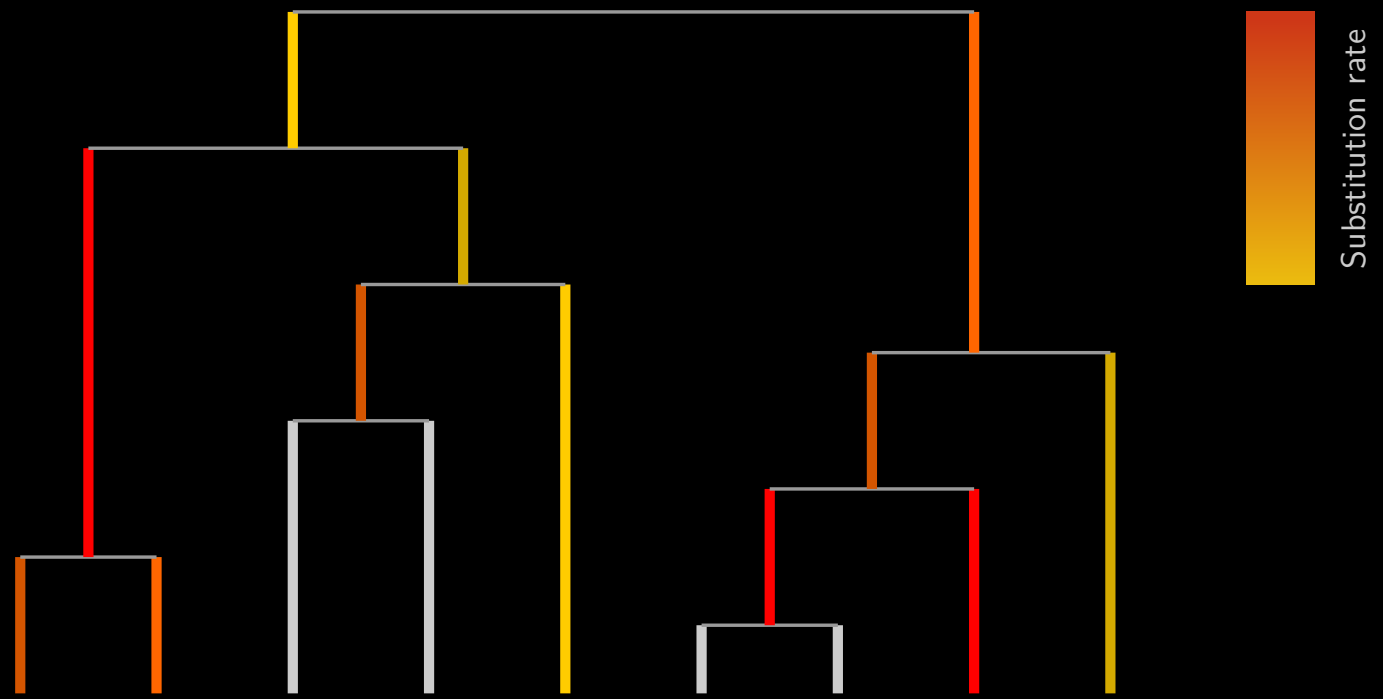
The non-auto-correlated ones first



- The substitution rate is different on each branch.
- Or the changing points are drawn first on the tree (Poisson process).
- *Rate values are chosen in a fixed law, independently of the neighbouring ones.*
- *Rate values depend on the rate in the parent branch.*

Non-auto-correlated v.s. auto-correlated relaxed clocks

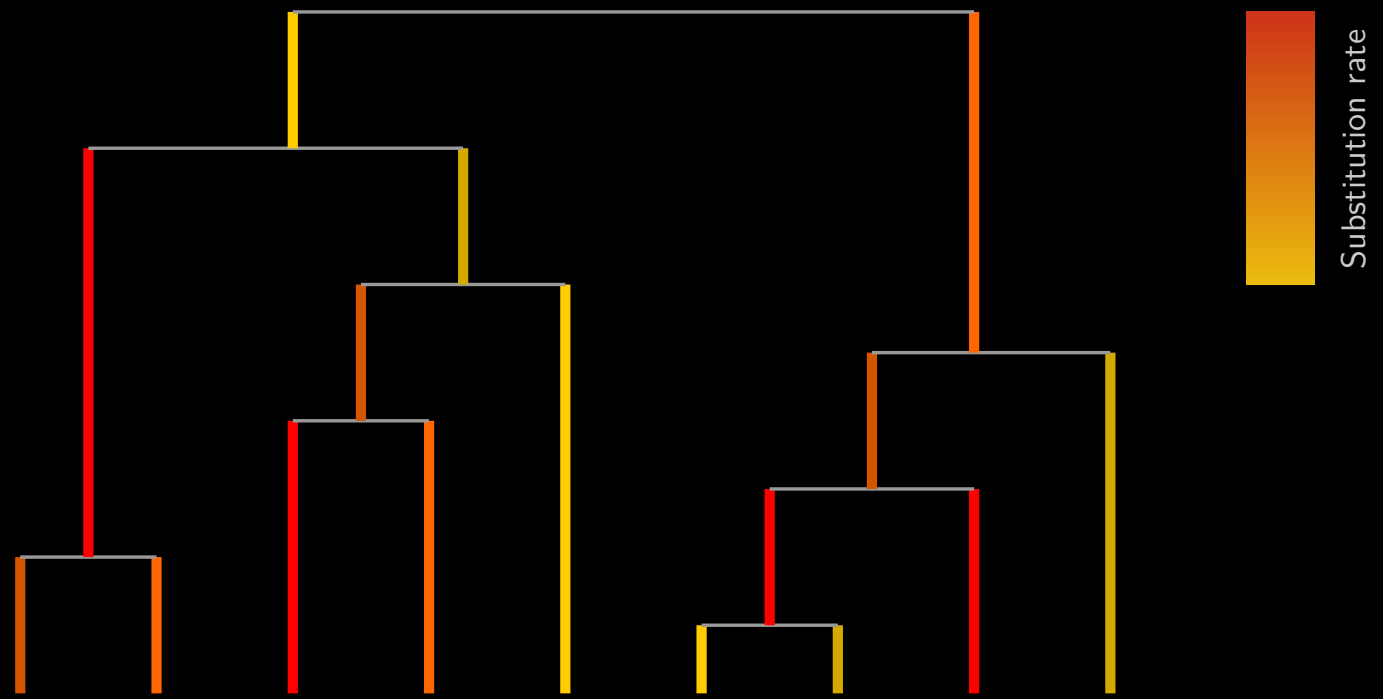
The non-auto-correlated ones first



- The substitution rate is different on each branch.
- Or the changing points are drawn first on the tree (Poisson process).
- *Rate values are chosen in a fixed law, independently of the neighbouring ones.*
- *Rate values depend on the rate in the parent branch.*

Non-auto-correlated v.s. auto-correlated relaxed clocks

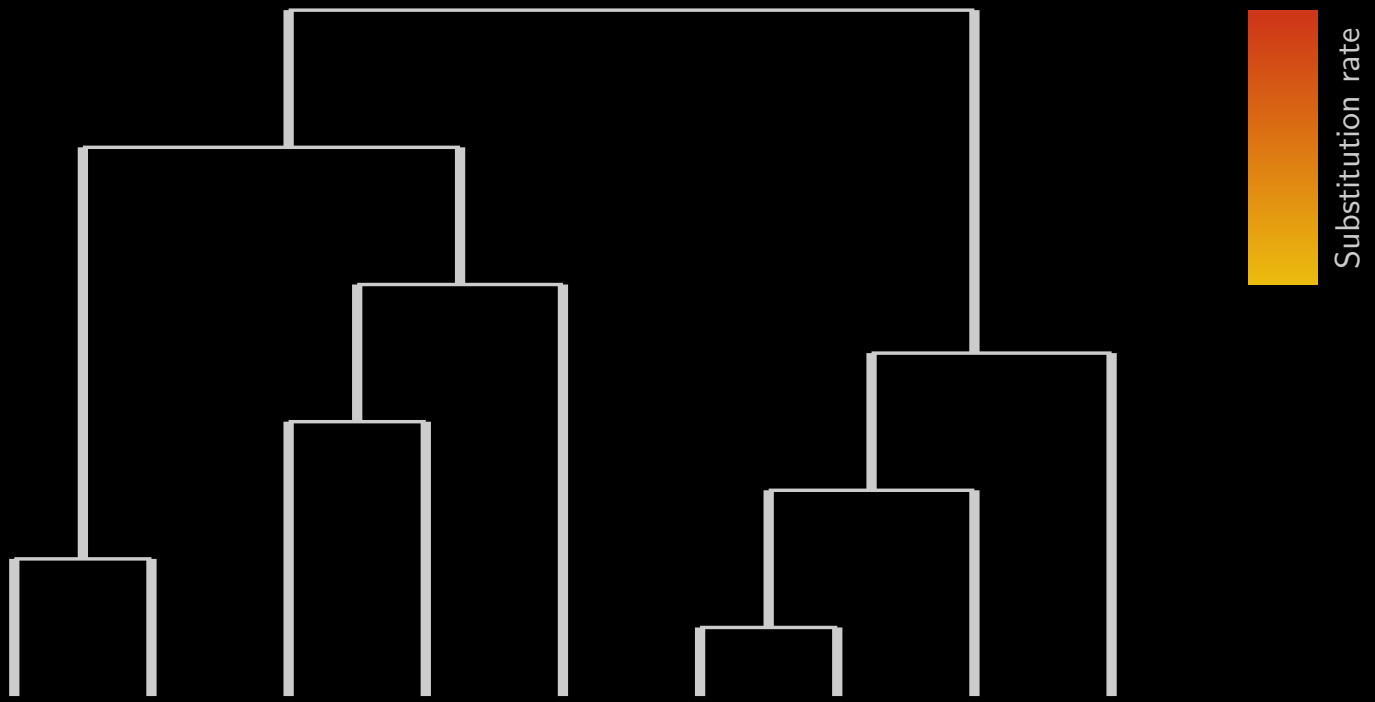
The non-auto-correlated ones first



- The substitution rate is different on each branch.
- Or the changing points are drawn first on the tree (Poisson process).
- *Rate values are chosen in a fixed law, independently of the neighbouring ones.*
- *Rate values depend on the rate in the parent branch.*

Non-auto-correlated v.s. auto-correlated relaxed clocks

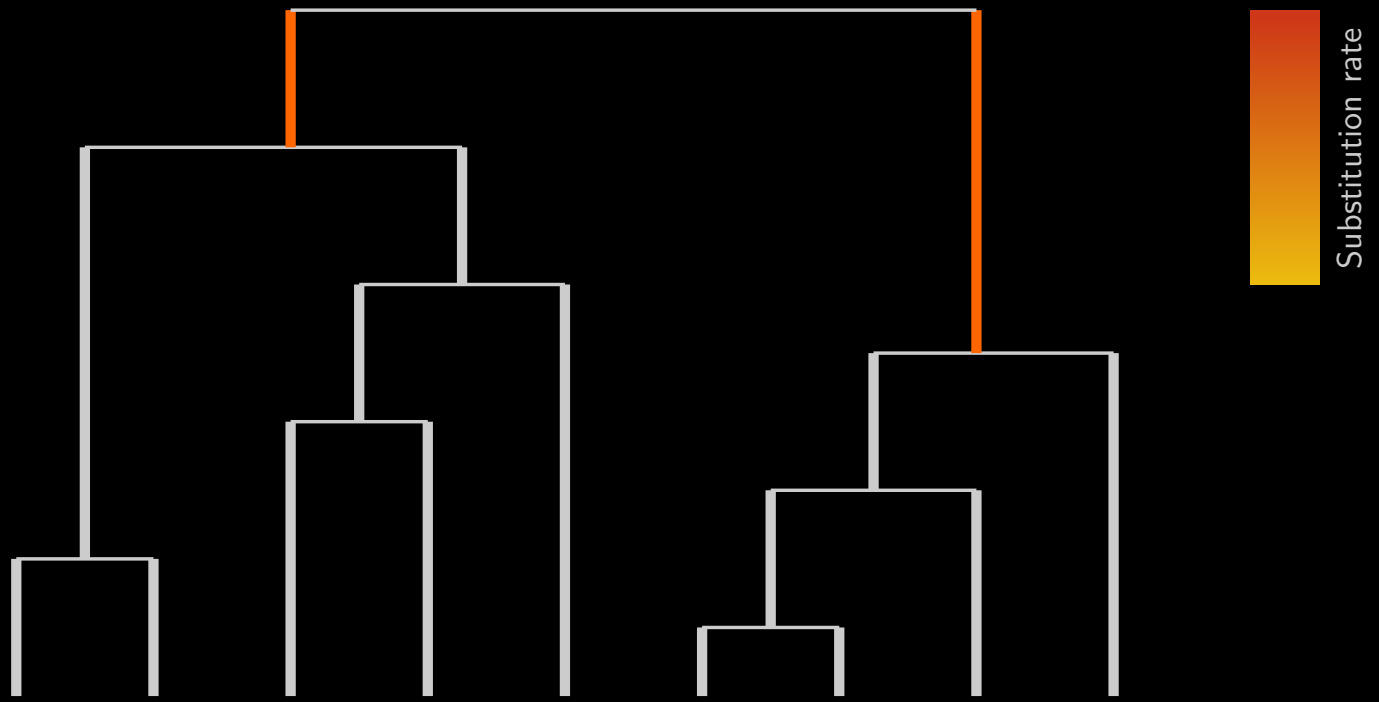
The auto-correlated ones



- The substitution rate is different on each branch.
- Or the changing points are drawn first on the tree (Poisson process).
- *Rate values are chosen in a fixed law, independently of the neighbouring ones.*
- *Rate values depend on the rate in the parent branch.*

Non-auto-correlated v.s. auto-correlated relaxed clocks

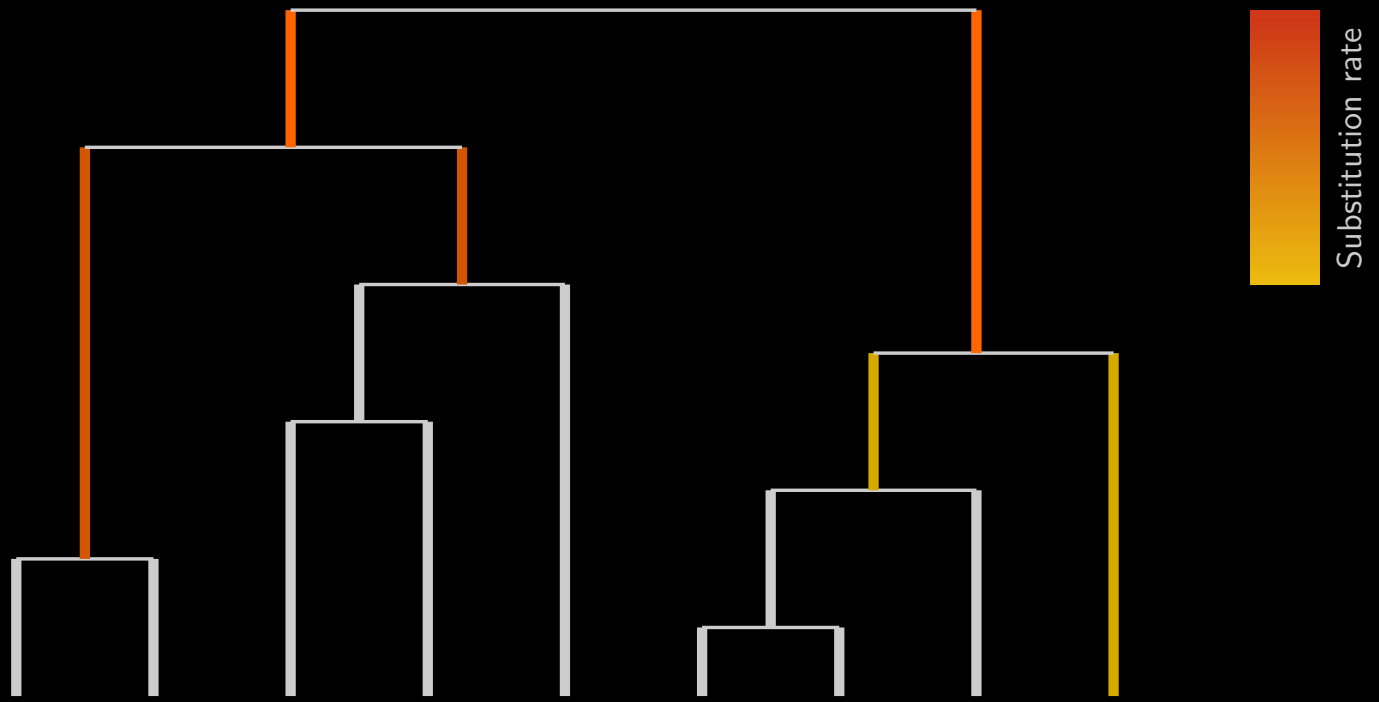
The auto-correlated ones



- The substitution rate is different on each branch.
- Or the changing points are drawn first on the tree (Poisson process).
- *Rate values are chosen in a fixed law, independently of the neighbouring ones.*
- *Rate values depend on the rate in the parent branch.*

Non-auto-correlated v.s. auto-correlated relaxed clocks

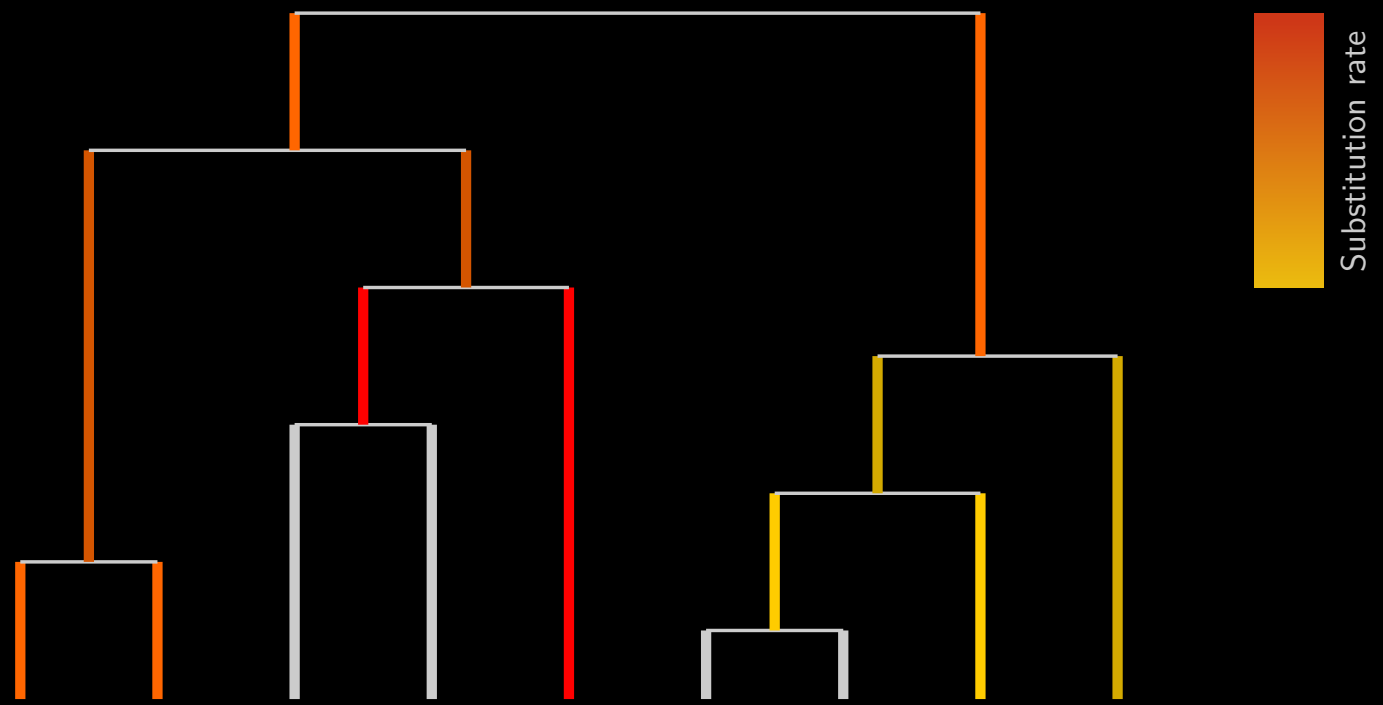
The auto-correlated ones



- The substitution rate is different on each branch.
- Or the changing points are drawn first on the tree (Poisson process).
- *Rate values are chosen in a fixed law, independently of the neighbouring ones.*
- *Rate values depend on the rate in the parent branch.*

Non-auto-correlated v.s. auto-correlated relaxed clocks

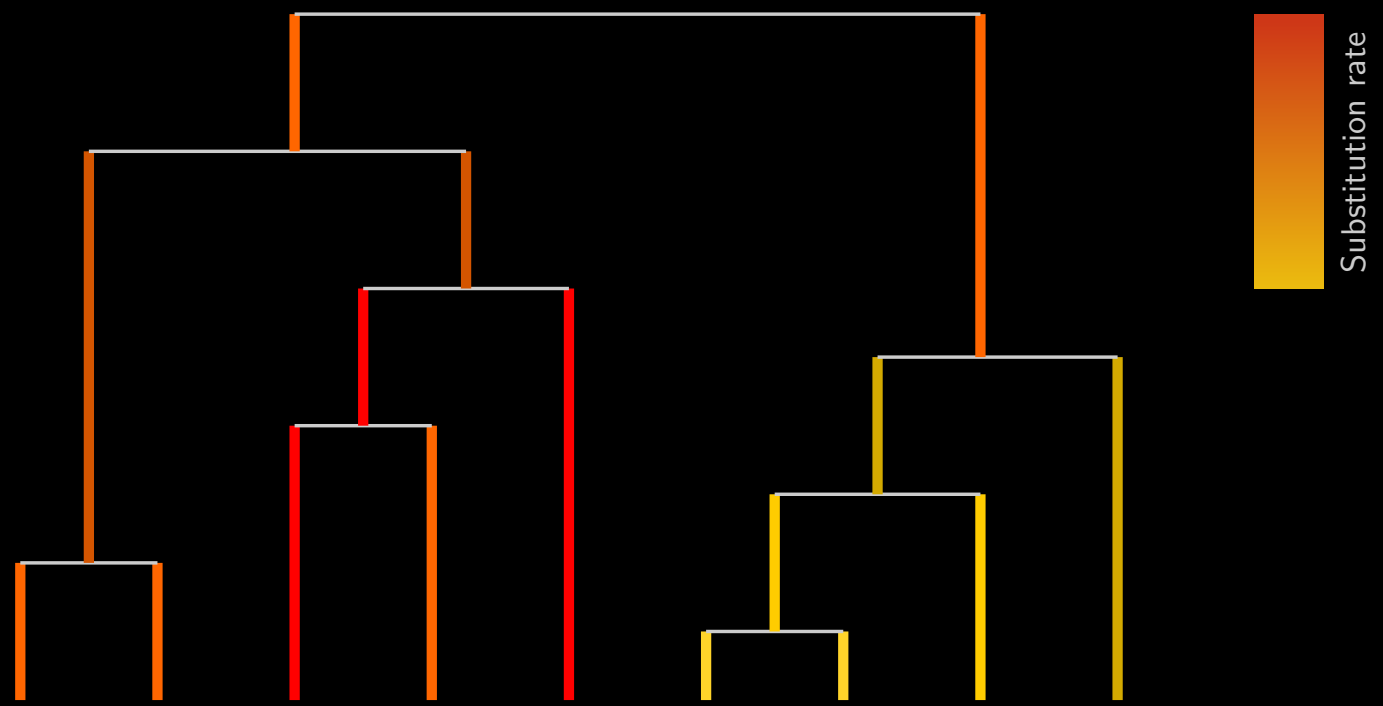
The auto-correlated ones



- The substitution rate is different on each branch.
- Or the changing points are drawn first on the tree (Poisson process).
- *Rate values are chosen in a fixed law, independently of the neighbouring ones.*
- *Rate values depend on the rate in the parent branch.*

Non-auto-correlated v.s. auto-correlated relaxed clocks

The auto-correlated ones



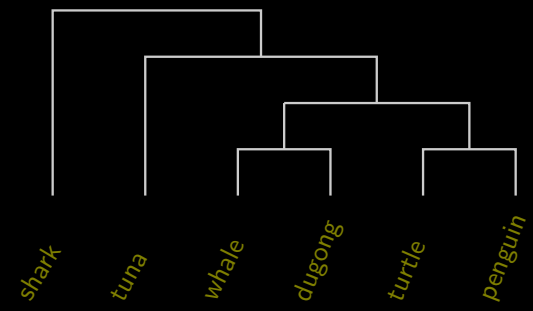
- The substitution rate is different on each branch.
- Or the changing points are drawn first on the tree (Poisson process).
- *Rate values are chosen in a fixed law, independently of the neighbouring ones.*
- *Rate values depend on the rate in the parent branch.*

The inference framework

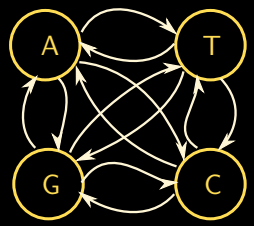
On a single locus first Lepage et al. (2007)

DATA

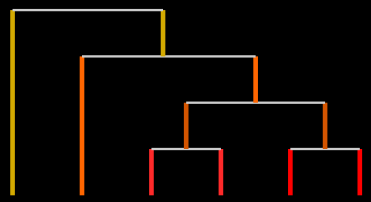
shark	A	T	C	G	A	A	G	C	...
tuna	-	-	G	-	C	T	-	-	...
whale	-	A	G	-	-	-	-	-	...
dugong	-	A	G	-	-	-	-	-	...
turtle	-	-	G	-	-	-	A	-	...
penguin	-	-	G	-	-	-	A	-	...



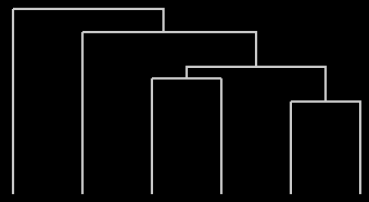
MODEL



$\mathbb{P}(\mathcal{A} \mid \mathcal{T}, \mathcal{R})$



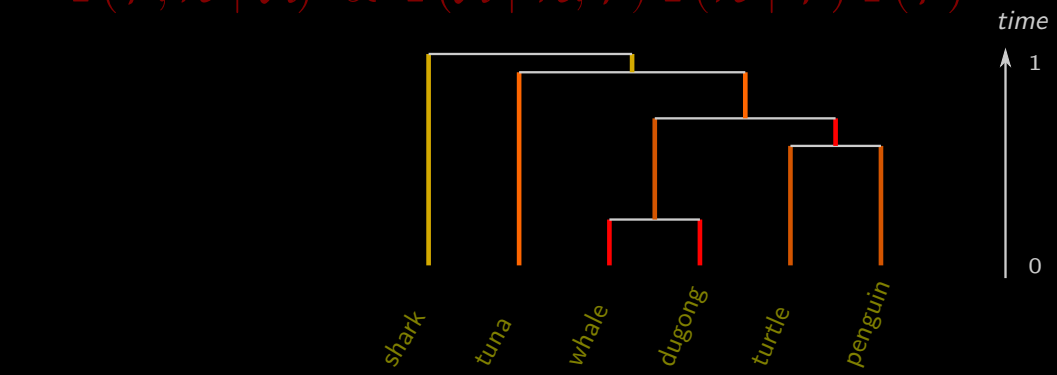
$\mathbb{P}(\mathcal{R} \mid \mathcal{T})$



$\mathbb{P}(\mathcal{T})$

$$\mathbb{P}(\mathcal{T}, \mathcal{R} \mid \mathcal{A}) \propto \mathbb{P}(\mathcal{A} \mid \mathcal{R}, \mathcal{T})\, \mathbb{P}(\mathcal{R} \mid \mathcal{T})\, \mathbb{P}(\mathcal{T})$$

OUTPUT

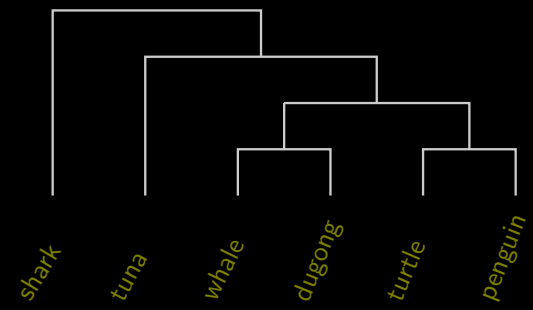


The inference framework

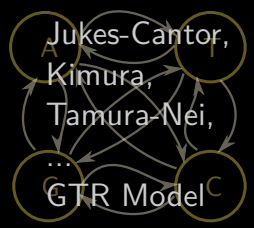
On a single locus first Lepage et al. (2007)

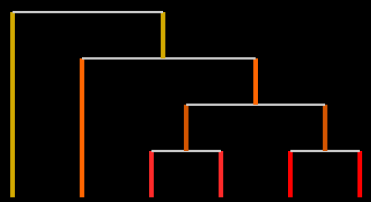
DATA

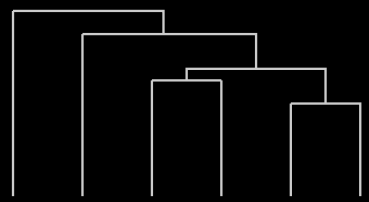
shark	A	T	C	G	A	A	G	C	...
tuna	-	-	G	-	C	T	-	-	...
whale	-	A	G	-	-	-	-	-	...
dugong	-	A	G	-	-	-	-	-	...
turtle	-	-	G	-	-	-	A	-	...
penguin	-	-	G	-	-	-	A	-	...



MODEL



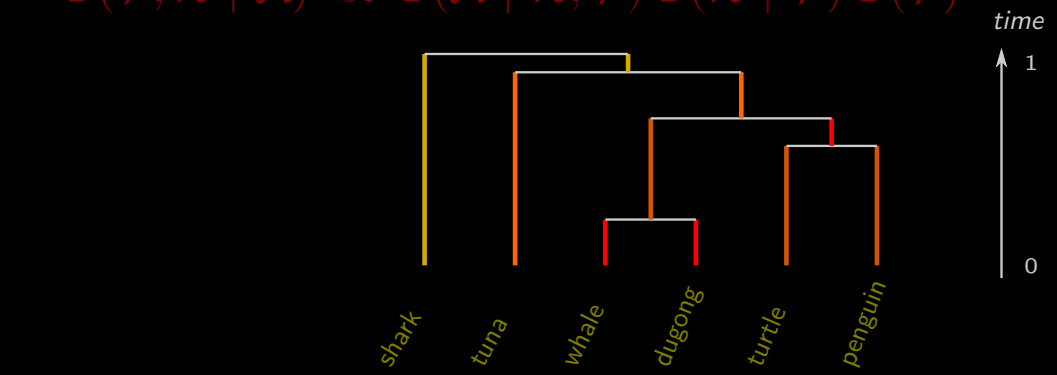




$\mathbb{P}(\mathcal{A} \mid \mathcal{T}, \mathcal{R})$ $\mathbb{P}(\mathcal{R} \mid \mathcal{T})$ $\mathbb{P}(\mathcal{T})$

$$\mathbb{P}(\mathcal{T}, \mathcal{R} \mid \mathcal{A}) \propto \mathbb{P}(\mathcal{A} \mid \mathcal{R}, \mathcal{T})\, \mathbb{P}(\mathcal{R} \mid \mathcal{T})\, \mathbb{P}(\mathcal{T})$$

OUTPUT

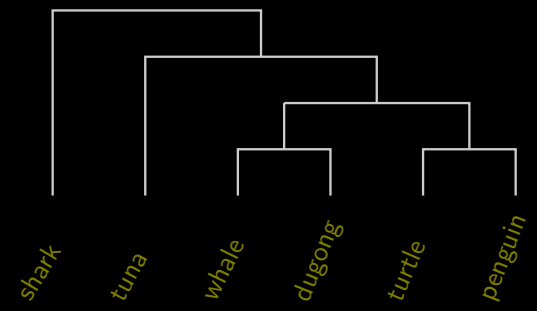


The inference framework

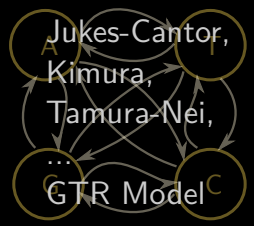
On a single locus first Lepage et al. (2007)

DATA

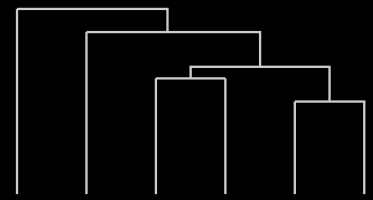
shark	A	T	C	G	A	A	G	C	...
tuna	-	-	G	-	C	T	-	-	...
whale	-	A	G	-	-	-	-	-	...
dugong	-	A	G	-	-	-	-	-	...
turtle	-	-	G	-	-	-	A	-	...
penguin	-	-	G	-	-	-	A	-	...



MODEL



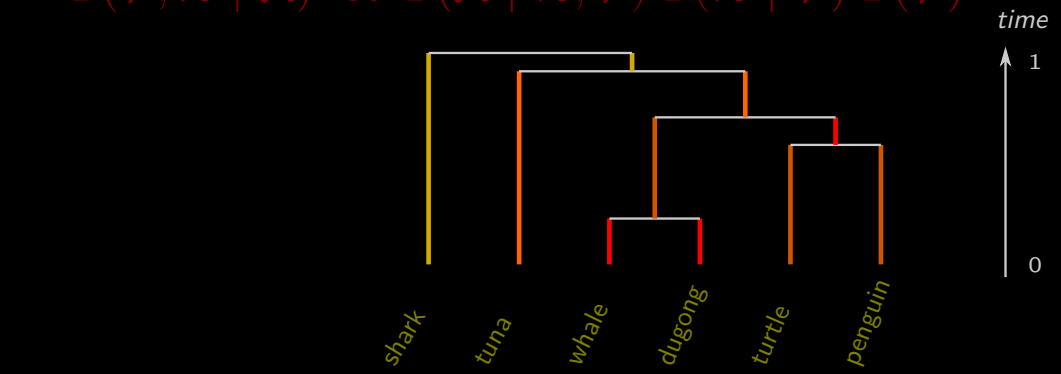
Non-auto-correlated models
Auto-correlated models
+
Choice of rate distribution



$\mathbb{P}(\mathcal{A} \mid \mathcal{T}, \mathcal{R})$ $\mathbb{P}(\mathcal{R} \mid \mathcal{T})$ $\mathbb{P}(\mathcal{T})$

$$\mathbb{P}(\mathcal{T}, \mathcal{R} \mid \mathcal{A}) \propto \mathbb{P}(\mathcal{A} \mid \mathcal{R}, \mathcal{T})\, \mathbb{P}(\mathcal{R} \mid \mathcal{T})\, \mathbb{P}(\mathcal{T})$$

OUTPUT

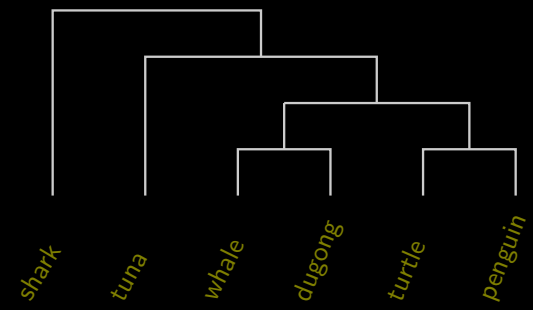


The inference framework

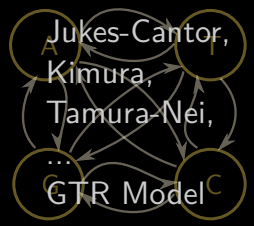
On a single locus first Lepage et al. (2007)

DATA

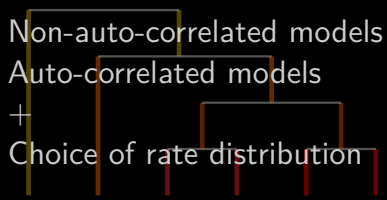
shark	A	T	C	G	A	A	G	C	...
tuna	-	-	G	-	C	T	-	-	...
whale	-	A	G	-	-	-	-	-	...
dugong	-	A	G	-	-	-	-	-	...
turtle	-	-	G	-	-	-	A	-	...
penguin	-	-	G	-	-	-	A	-	...



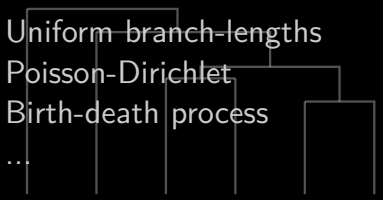
MODEL



Jukes-Cantor,
Kimura,
Tamura-Nei,
...
GTR Model



Non-auto-correlated models
Auto-correlated models
+
Choice of rate distribution

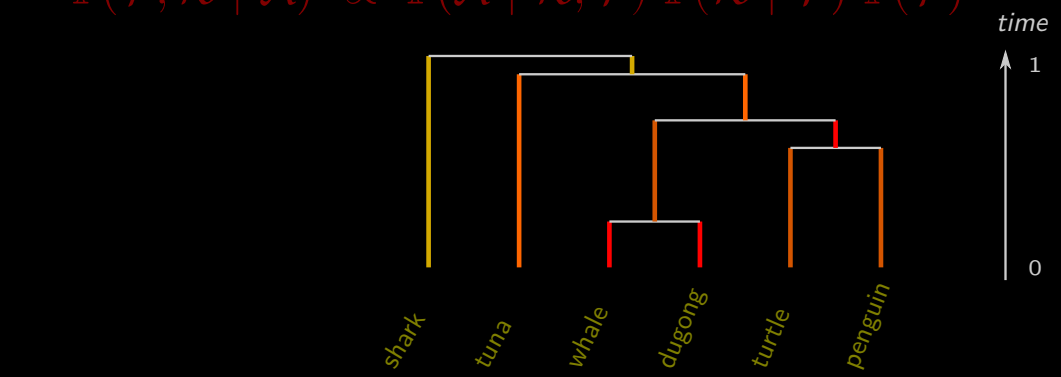


Uniform branch-lengths
Poisson-Dirichlet
Birth-death process
...

$\mathbb{P}(\mathcal{A} \mid \mathcal{T}, \mathcal{R})$ $\mathbb{P}(\mathcal{R} \mid \mathcal{T})$ $\mathbb{P}(\mathcal{T})$

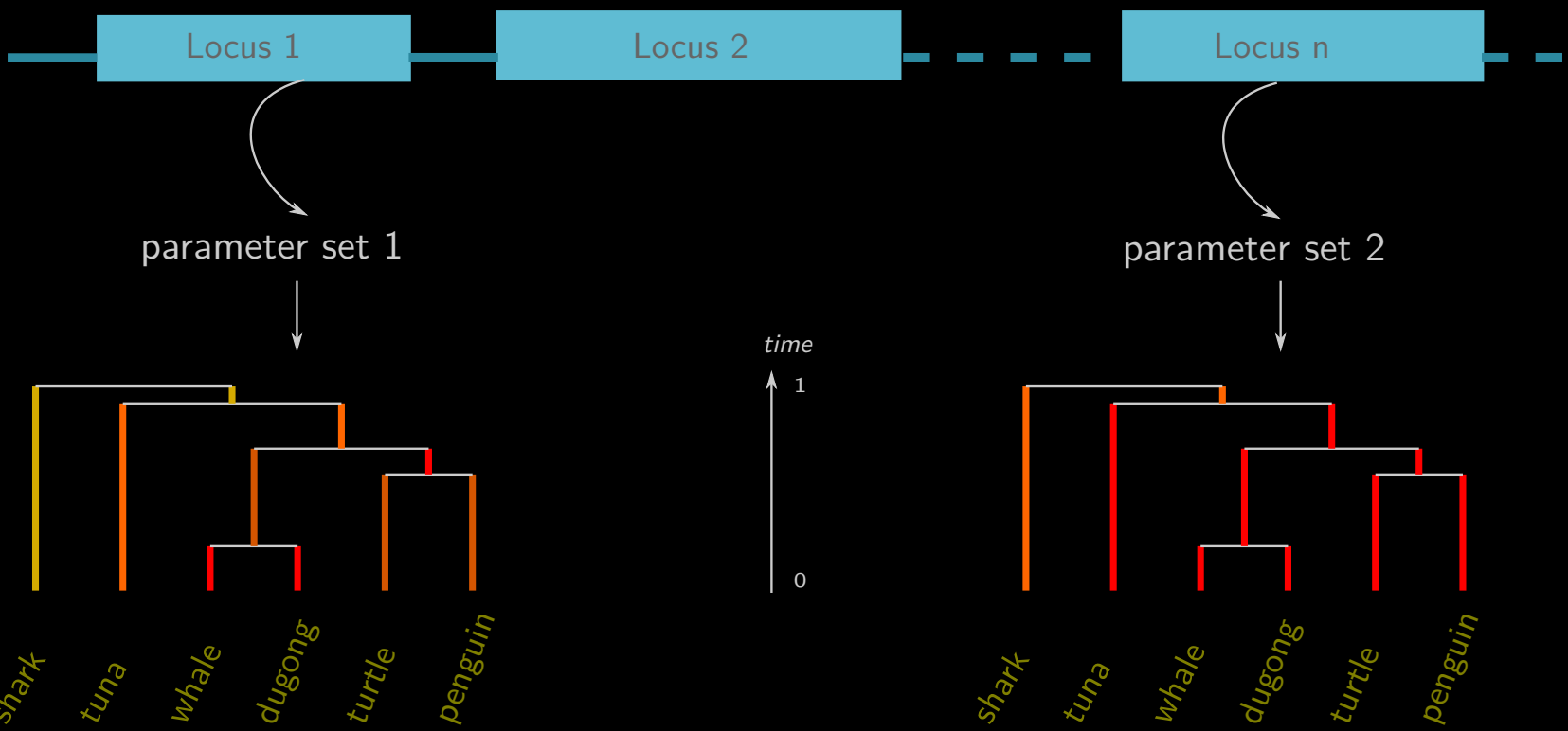
$$\mathbb{P}(\mathcal{T}, \mathcal{R} \mid \mathcal{A}) \propto \mathbb{P}(\mathcal{A} \mid \mathcal{R}, \mathcal{T})\, \mathbb{P}(\mathcal{R} \mid \mathcal{T})\, \mathbb{P}(\mathcal{T})$$

OUTPUT



The inference framework

Along the sequence



- Model parameters may change from one locus to the other.
- Each parameter set is drawn in a given distribution.
- Need for a prior on the partition of loci behaving similarly.

Progress for our understanding of evolution

More accuracy in dating

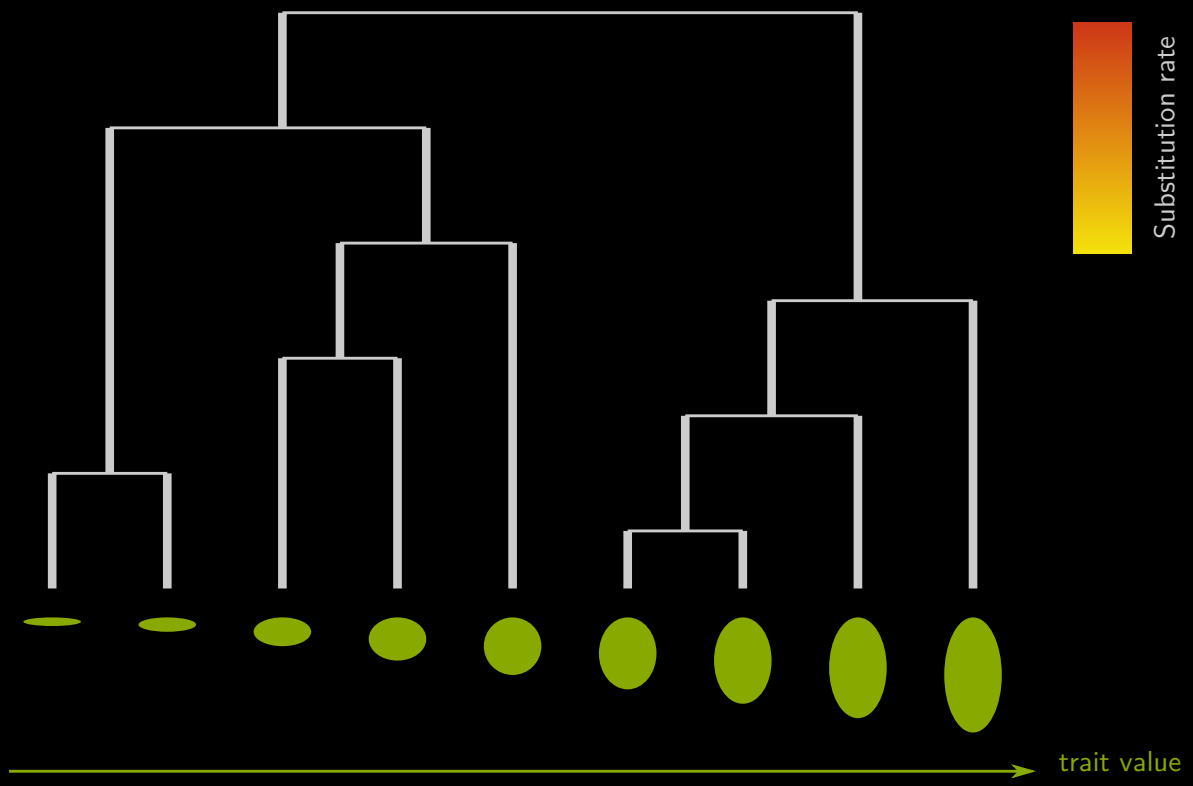
- Three ways to evaluate these clocks :

1. Ability to retrieve known parameter values on simulations.
2. Goodness of fit on specific datasets.
3. Compatibility with known fossil dates.

- Relaxed clocks perform better than the strict one on most datasets.
- Some studies compare the distinct relaxed clocks (Lepage et al., 2007).

Progress for our understanding of evolution

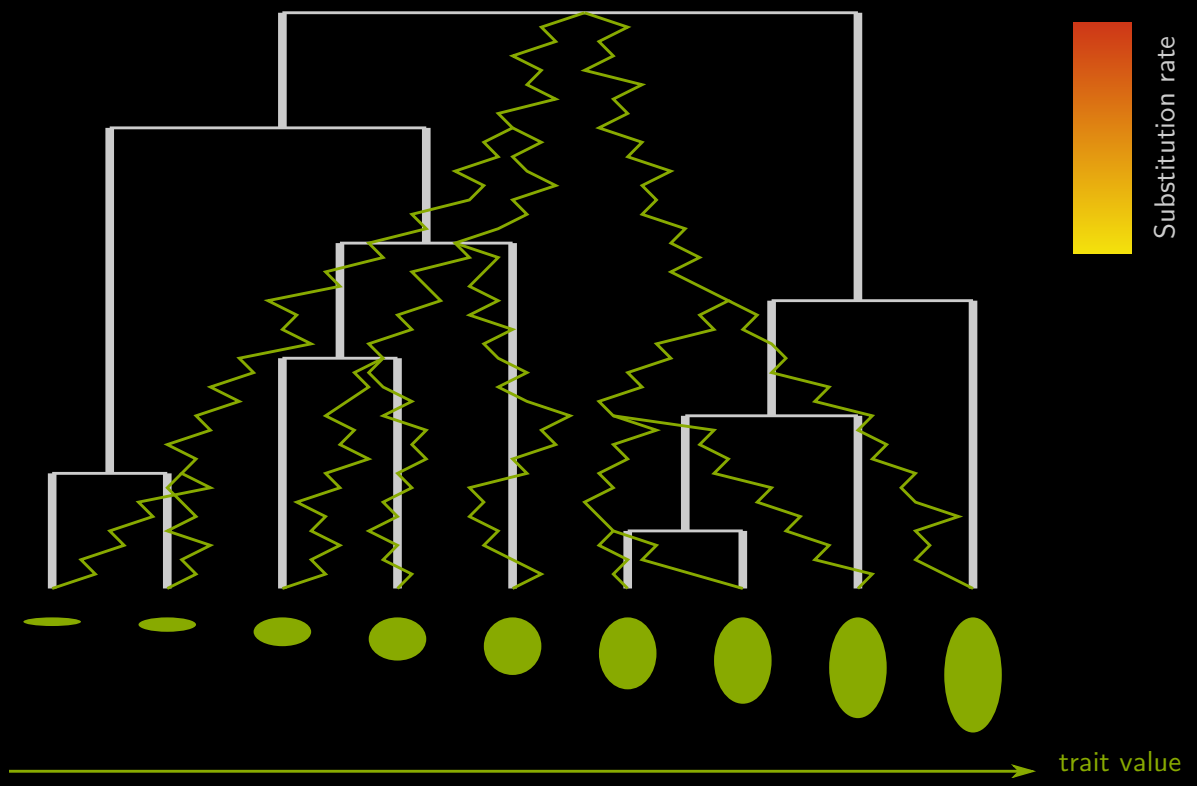
Correlation with phenotypic traits (Lartillot and Poujol, 2011)



- A continuous trait (like body mass) changes continuously along tree branches.
- The rate of substitution changes continuously along tree branches.
- Both are modeled jointly as a correlated diffusion process.

Progress for our understanding of evolution

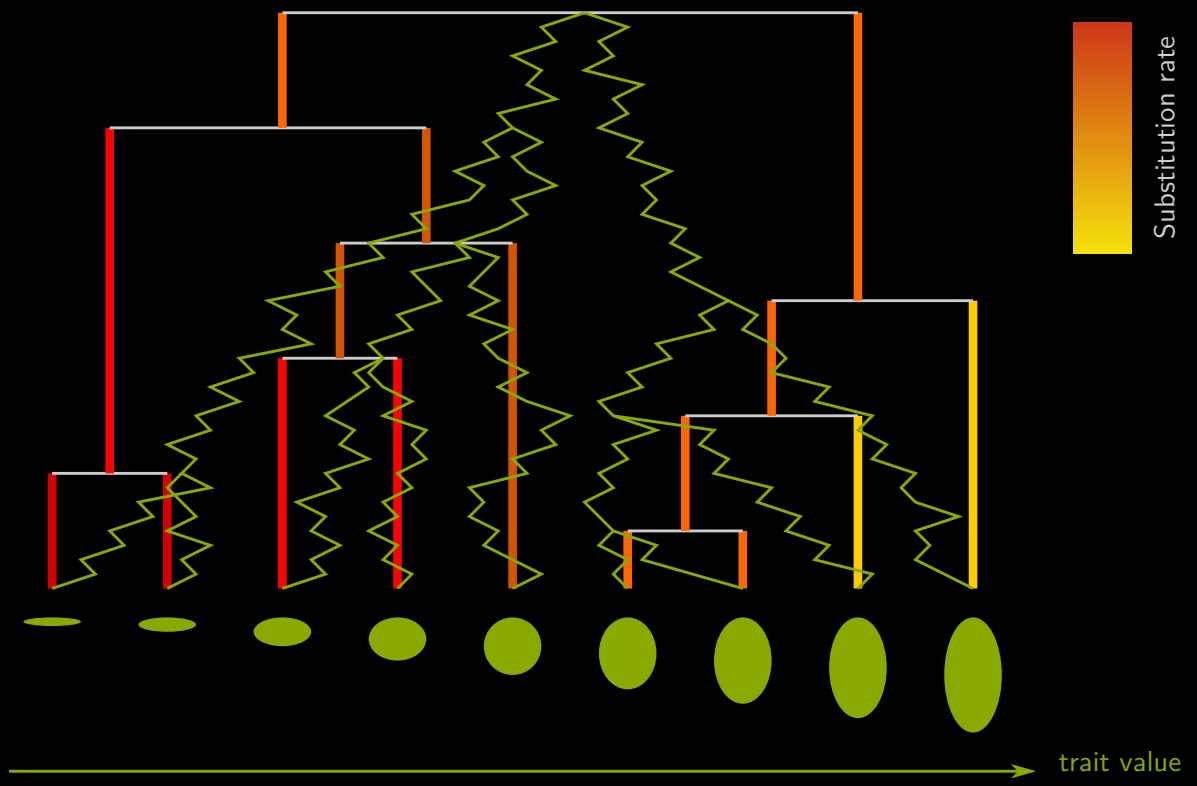
Correlation with phenotypic traits (Lartillot and Poujol, 2011)



- A continuous trait (like body mass) changes continuously along tree branches.
- The rate of substitution changes continuously along tree branches.
- Both are modeled jointly as a correlated diffusion process.

Progress for our understanding of evolution

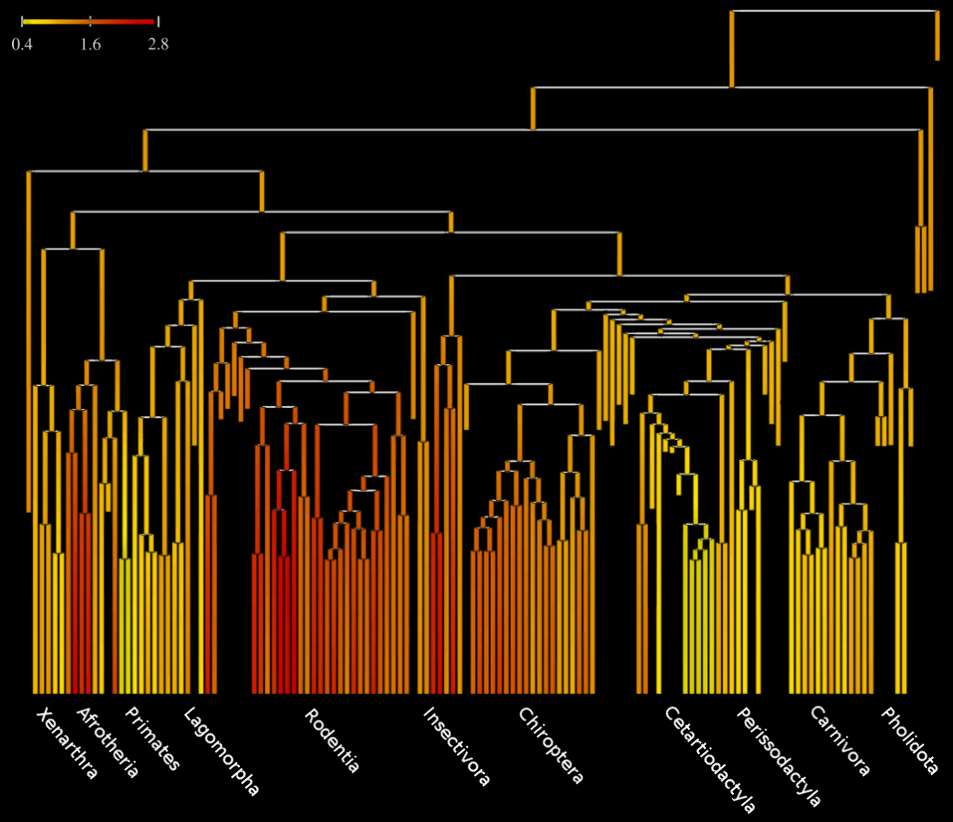
Correlation with phenotypic traits (Lartillot and Poujol, 2011)



- A continuous trait (like body mass) changes continuously along tree branches.
- The rate of substitution changes continuously along tree branches.
- Both are modeled jointly as a correlated diffusion process.

Progress for our understanding of evolution

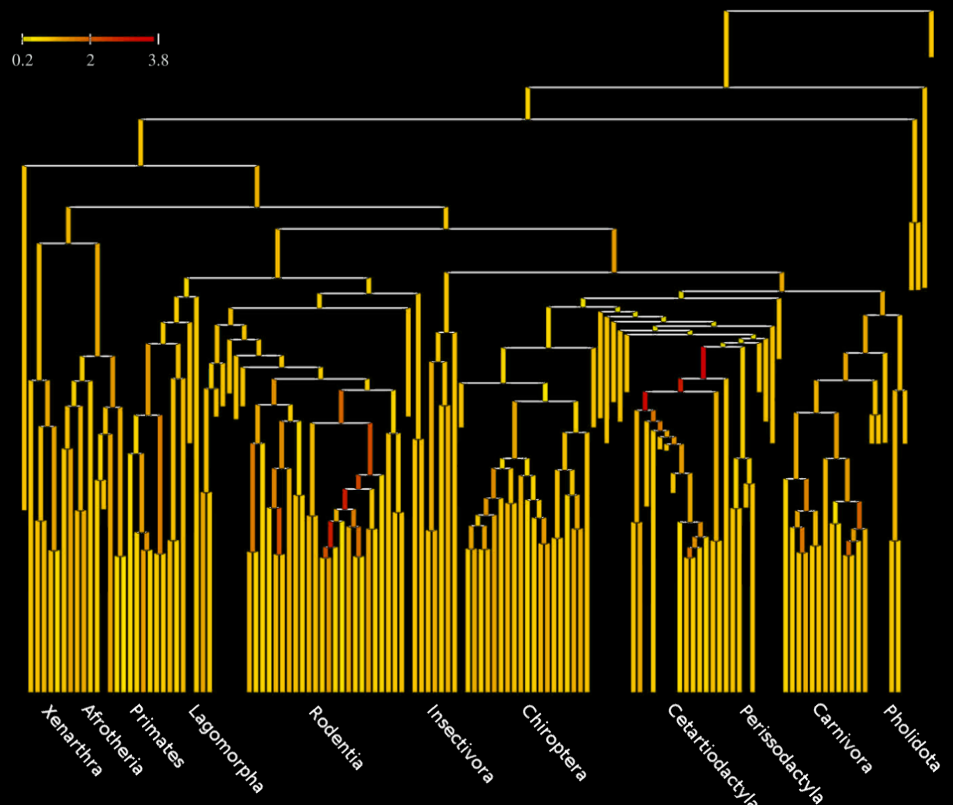
Estimating the weight of the non-auto-correlated v.s. auto-correlated parts (Lartillot et al., 2016)



- Among mammals, there is a negative correlation between :
  1. substitution rate and body mass,
  2. substitution rate and longevity.
- *Which processes are subsumed under the 'non-auto-correlated' term ?*

Progress for our understanding of evolution

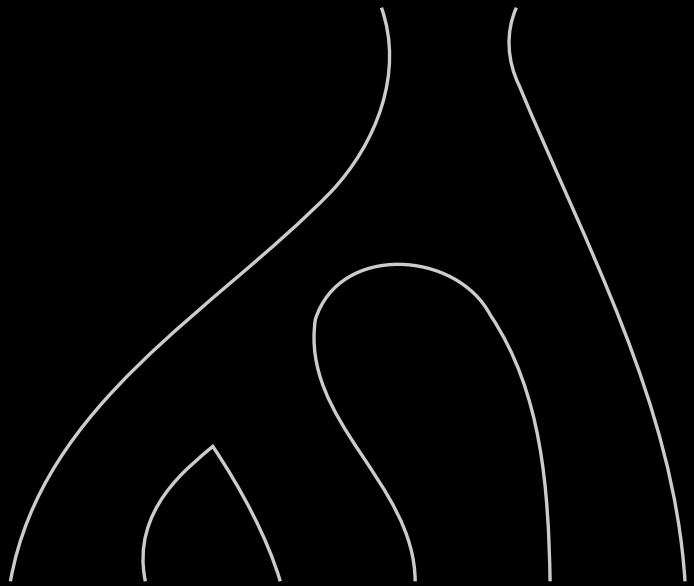
Estimating the weight of the non-auto-correlated v.s. auto-correlated parts (Lartillot et al., 2016)



- Among mammals, there is a negative correlation between :
  1. substitution rate and body mass,
  2. substitution rate and longevity.
- *Which processes are subsumed under the 'non-auto-correlated' term ?*

Relevant biological knowledge to inform relaxed clocks

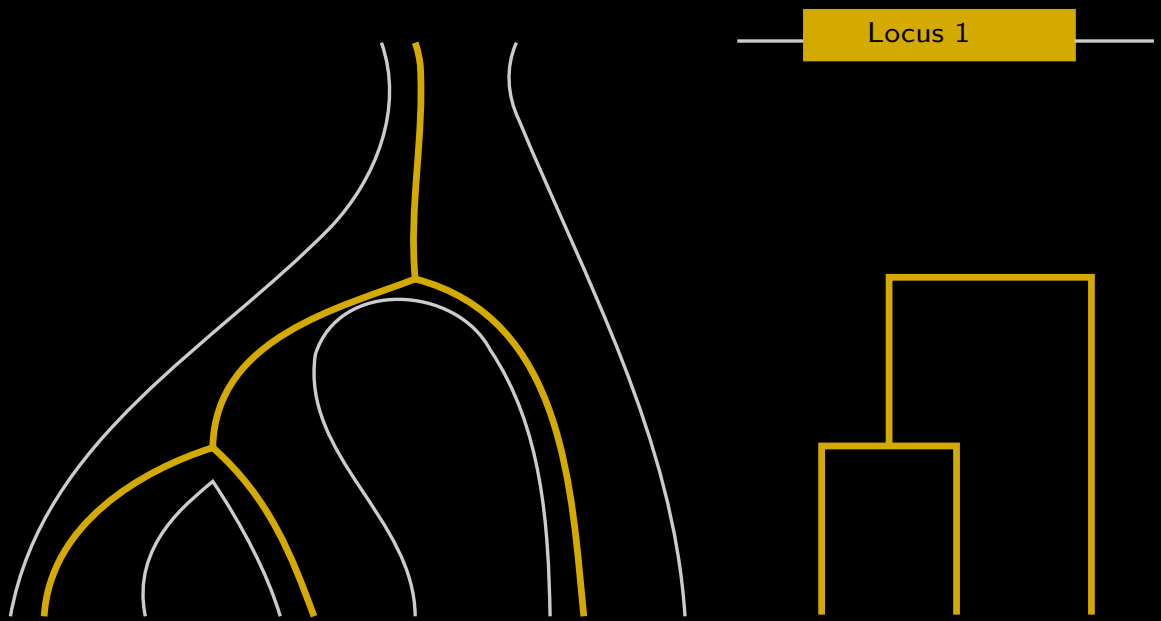
Gene coalescence depth



- Coalescence times vary randomly from one locus to the other.
  This should not be interpreted as a variation of the substitution rate.
  Could we inform the model with a coalescence-compatible time window ?
- Genes involved in reproductive isolation are expected to coalesce earlier.
  Could we look for such quickly coalescing genes ?

Relevant biological knowledge to inform relaxed clocks

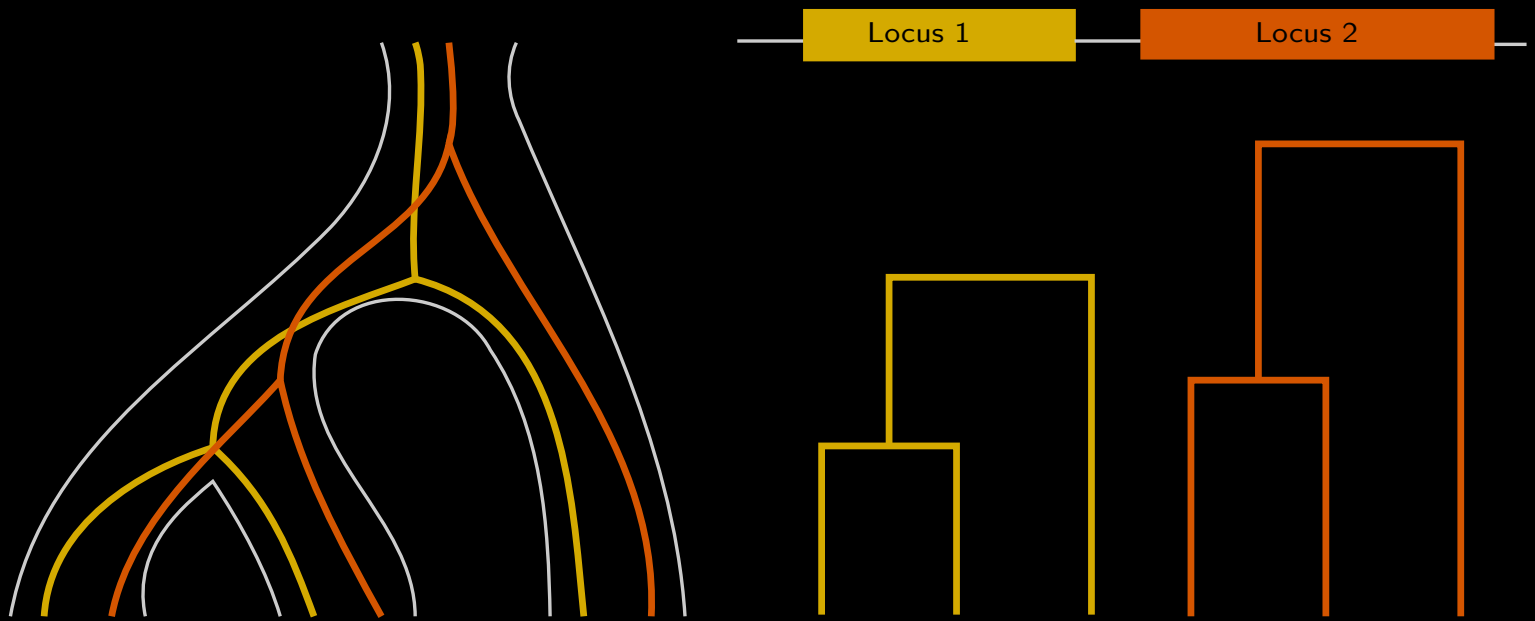
Gene coalescence depth



- Coalescence times vary randomly from one locus to the other.
  This should not be interpreted as a variation of the substitution rate.
  Could we inform the model with a coalescence-compatible time window ?
- Genes involved in reproductive isolation are expected to coalesce earlier.
  Could we look for such quickly coalescing genes ?

Relevant biological knowledge to inform relaxed clocks

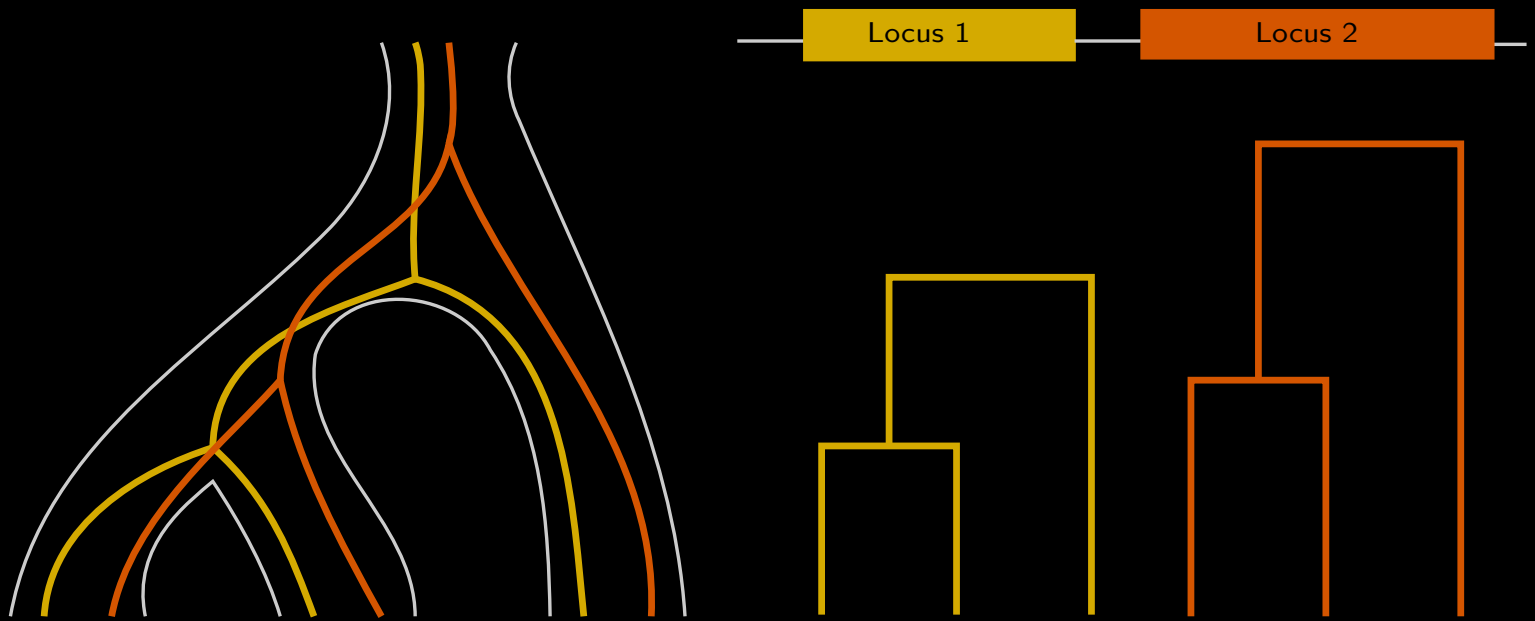
Gene coalescence depth



- Coalescence times vary randomly from one locus to the other.
  This should not be interpreted as a variation of the substitution rate.
  Could we inform the model with a coalescence-compatible time window ?
- Genes involved in reproductive isolation are expected to coalesce earlier.
  Could we look for such quickly coalescing genes ?

Relevant biological knowledge to inform relaxed clocks

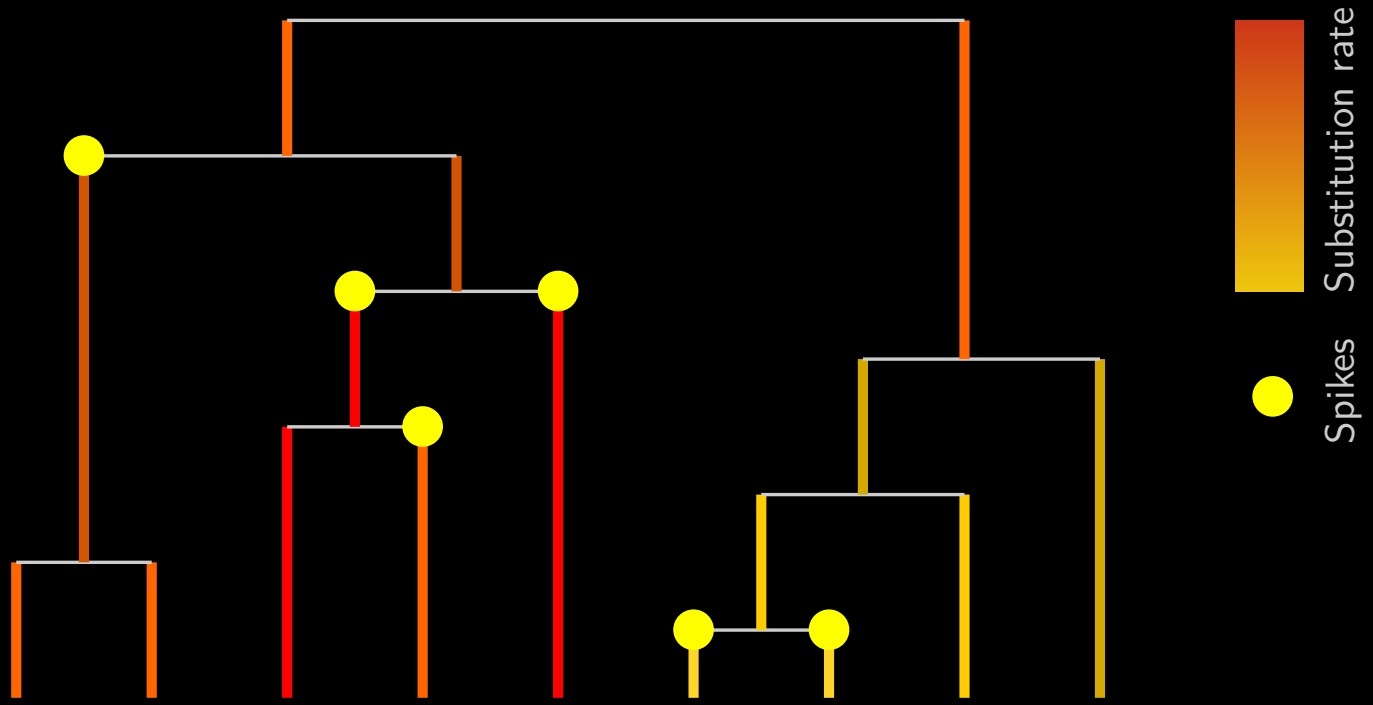
Gene coalescence depth



- Coalescence times vary randomly from one locus to the other.
  This should not be interpreted as a variation of the substitution rate.
  Could we inform the model with a coalescence-compatible time window ?
- Genes involved in reproductive isolation are expected to coalesce earlier.
  Could we look for such quickly coalescing genes ?

Relevant biological knowledge to inform relaxed clocks

Fast divergent selection at speciation times

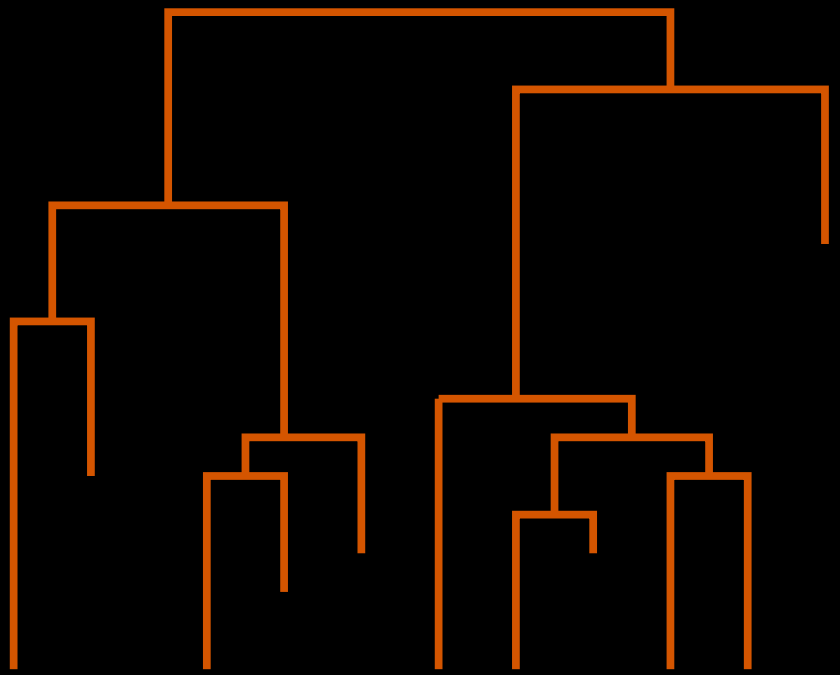


- Strong divergent selection is expected to happen at speciation.
- Concerns genes involved in ressource use, new habitat, conspecific recognition...
- *Could we look for these genes ?*

Looking specifically for 'ecologically diverging genes'

Positionning the spikes on the tree

From the whole process...

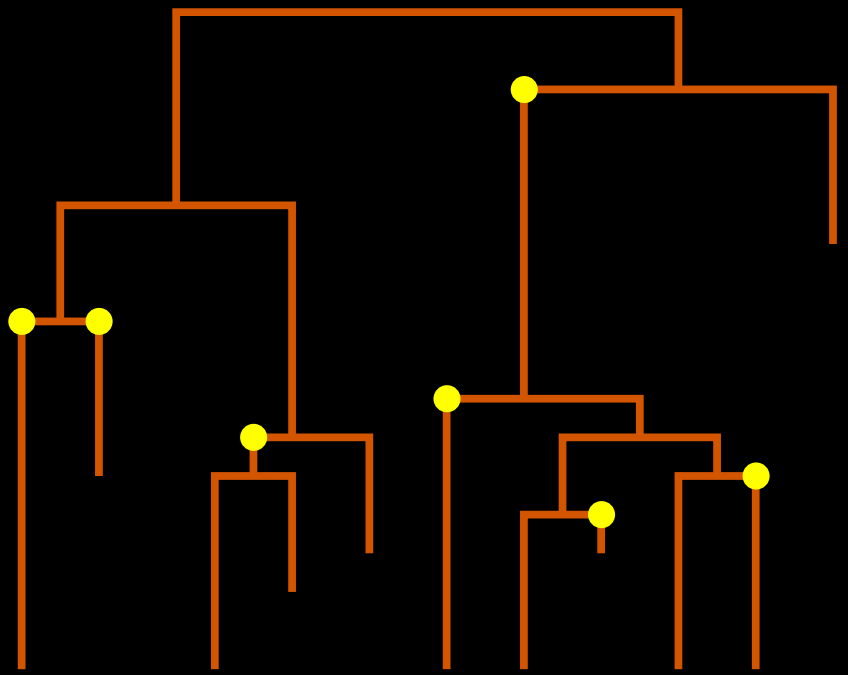


- The tree follows a birth-death process.
- At each branching time, a spike occurs with probability ν.
- We want to describe the process directly on the reconstructed tree.
- Spikes occur on the reconstructed tree as a time-heterogeneous Poisson process.

Looking specifically for 'ecologically diverging genes'

Positionning the spikes on the tree

From the whole process...

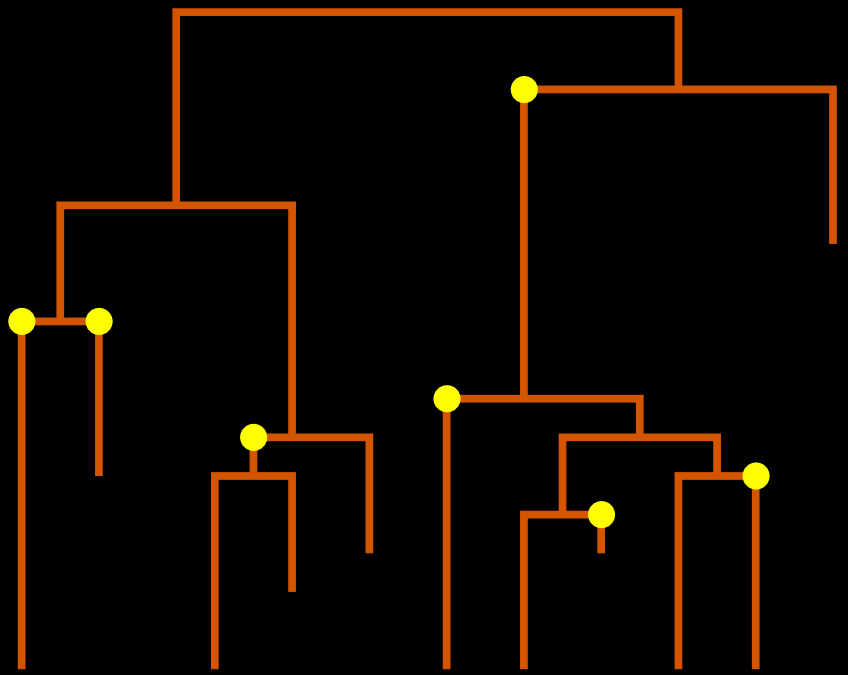


- The tree follows a birth-death process.
- At each branching time, a spike occurs with probability ν.
- We want to describe the process directly on the reconstructed tree.
- Spikes occur on the reconstructed tree as a time-heterogeneous Poisson process.

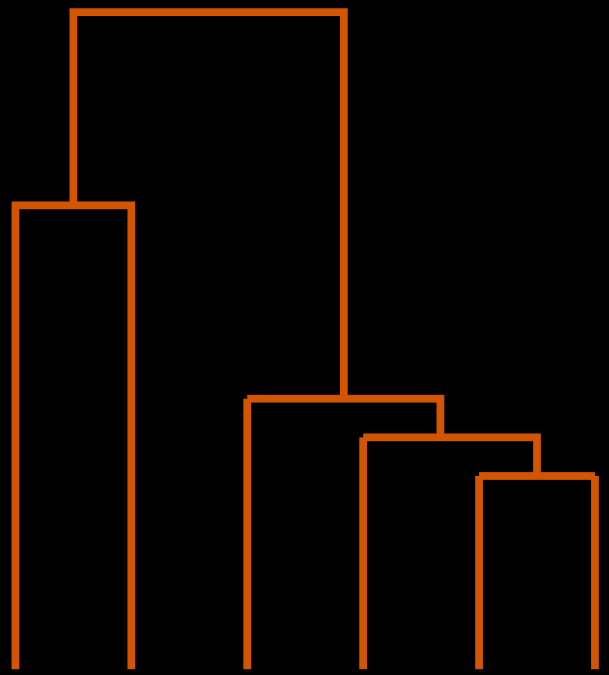
Looking specifically for 'ecologically diverging genes'

Positionning the spikes on the tree

From the whole process...



...to the reconstructed tree.

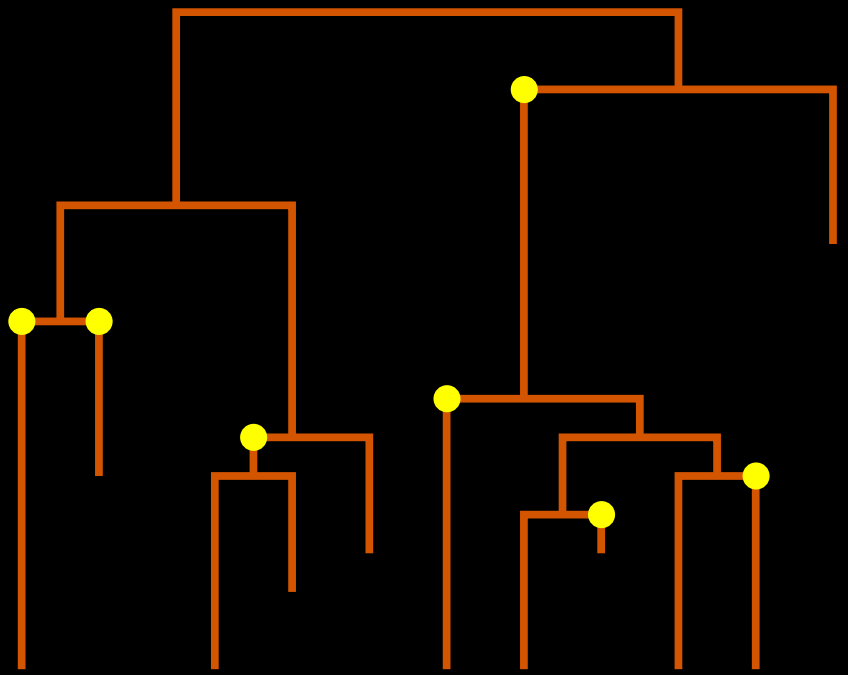


- The tree follows a birth-death process.
- At each branching time, a spike occurs with probability ν.
- We want to describe the process directly on the reconstructed tree.
- Spikes occur on the reconstructed tree as a time-heterogeneous Poisson process.

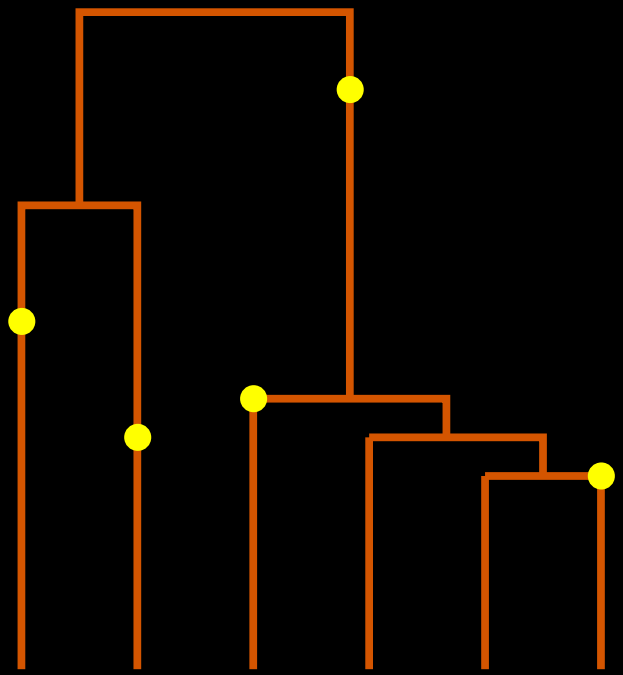
Looking specifically for 'ecologically diverging genes'

Positionning the spikes on the tree

From the whole process...



...to the reconstructed tree.



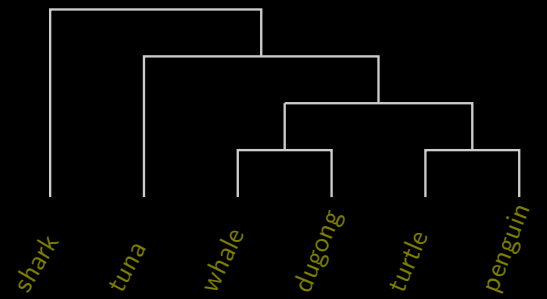
- The tree follows a birth-death process.
- At each branching time, a spike occurs with probability ν.
- We want to describe the process directly on the reconstructed tree.
- Spikes occur on the reconstructed tree as a time-heterogeneous Poisson process.

Looking specifically for ‘ecologically diverging genes’

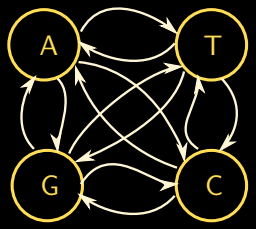
Same kind of framework

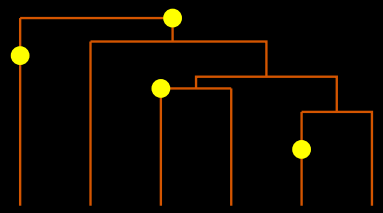
DATA

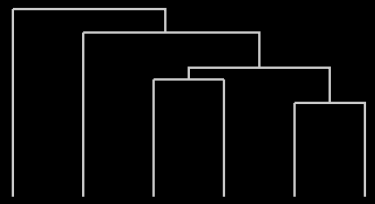
shark	A	T	C	G	A	A	G	C	...
tuna	-	-	G	-	C	T	-	-	...
whale	-	A	G	-	-	-	-	-	...
dugong	-	A	G	-	-	-	-	-	...
turtle	-	-	G	-	-	-	A	-	...
penguin	-	-	G	-	-	-	A	-	...



MODEL



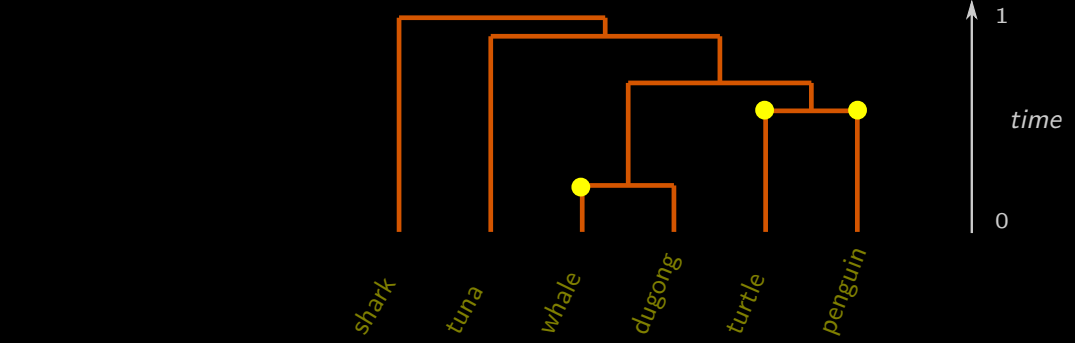




$\mathbb{P}(\mathcal{A} \mid \mathcal{T}, \mathcal{R})$ $\mathbb{P}(\mathcal{S} \mid \mathcal{T})$ $\mathbb{P}(\mathcal{T})$

$$\mathbb{P}(\mathcal{T}, \mathcal{S} \mid \mathcal{A}) \propto \mathbb{P}(\mathcal{A} \mid \mathcal{T}, \mathcal{S}) \mathbb{P}(\mathcal{S} \mid \mathcal{T}) \mathbb{P}(\mathcal{T})$$

OUTPUT

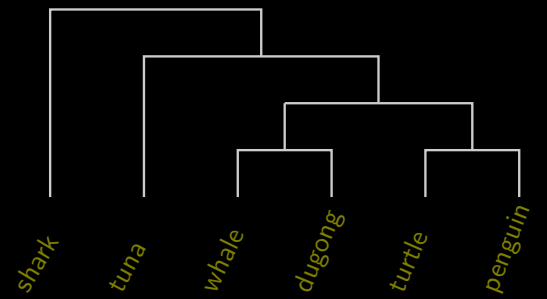


Looking specifically for 'ecologically diverging genes'

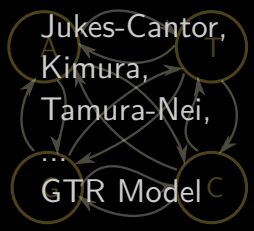
Same kind of framework

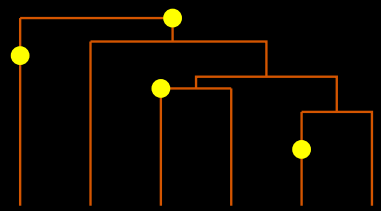
DATA

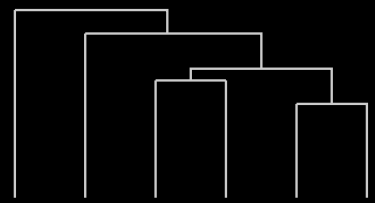
shark	A	T	C	G	A	A	G	C	...
tuna	-	-	G	-	C	T	-	-	...
whale	-	A	G	-	-	-	-	-	...
dugong	-	A	G	-	-	-	-	-	...
turtle	-	-	G	-	-	-	A	-	...
penguin	-	-	G	-	-	-	A	-	...



MODEL



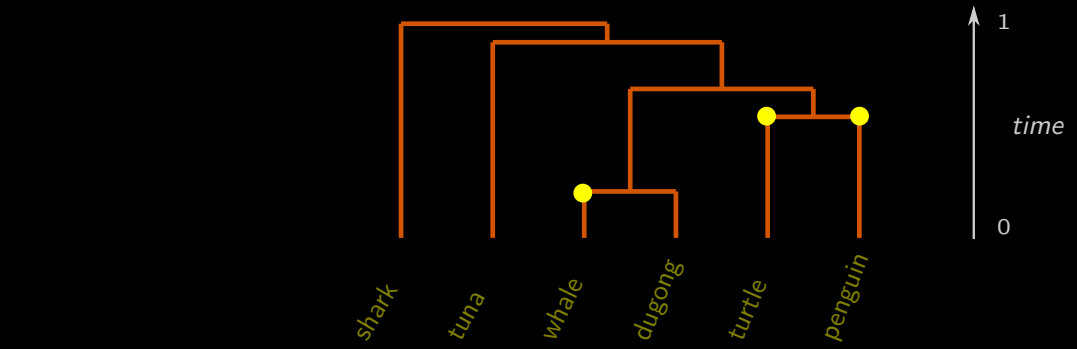




$\mathbb{P}(\mathcal{A} \mid \mathcal{T}, \mathcal{R})$ $\mathbb{P}(\mathcal{S} \mid \mathcal{T})$ $\mathbb{P}(\mathcal{T})$

$$\mathbb{P}(\mathcal{T}, \mathcal{S} \mid \mathcal{A}) \propto \mathbb{P}(\mathcal{A} \mid \mathcal{T}, \mathcal{S}) \mathbb{P}(\mathcal{S} \mid \mathcal{T}) \mathbb{P}(\mathcal{T})$$

OUTPUT

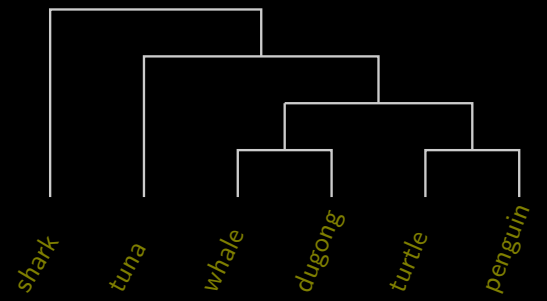


Looking specifically for 'ecologically diverging genes'

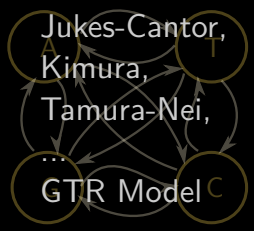
Same kind of framework

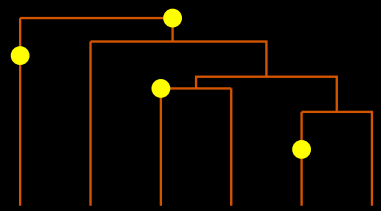
DATA

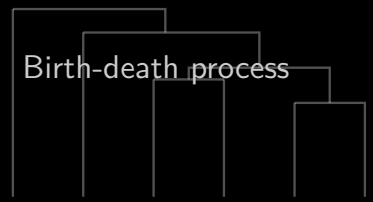
shark	A	T	C	G	A	A	G	C	...
tuna	-	-	G	-	C	T	-	-	...
whale	-	A	G	-	-	-	-	-	...
dugong	-	A	G	-	-	-	-	-	...
turtle	-	-	G	-	-	-	A	-	...
penguin	-	-	G	-	-	-	A	-	...



MODEL



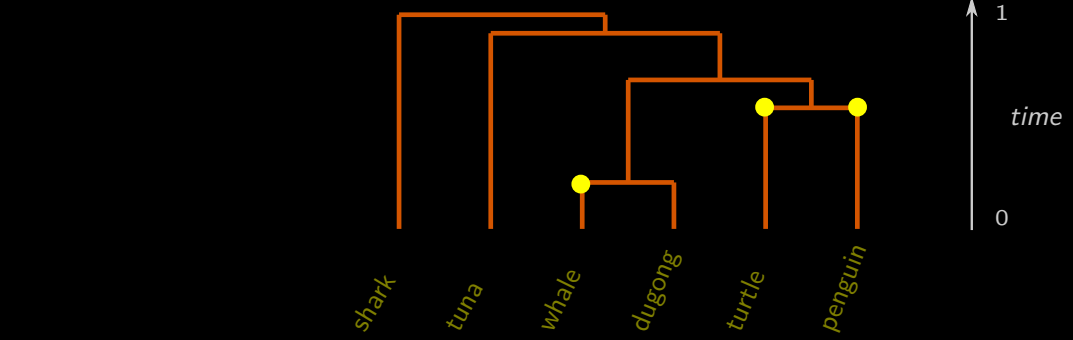




$\mathbb{P}(\mathcal{A} \mid \mathcal{T}, \mathcal{R})$ $\mathbb{P}(\mathcal{S} \mid \mathcal{T})$ $\mathbb{P}(\mathcal{T})$

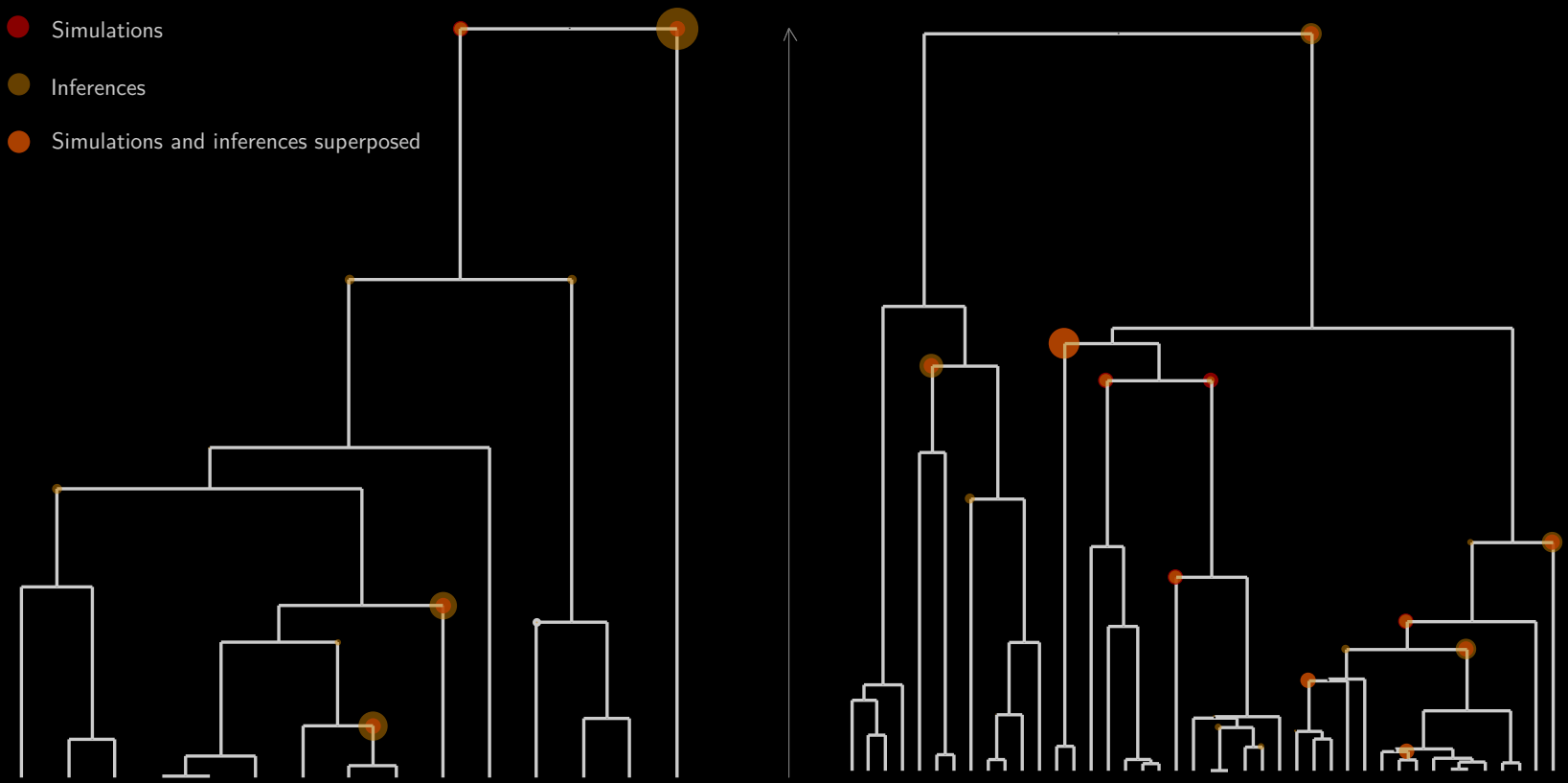
$$\mathbb{P}(\mathcal{T}, \mathcal{S} \mid \mathcal{A}) \propto \mathbb{P}(\mathcal{A} \mid \mathcal{T}, \mathcal{S}) \mathbb{P}(\mathcal{S} \mid \mathcal{T}) \mathbb{P}(\mathcal{T})$$

OUTPUT



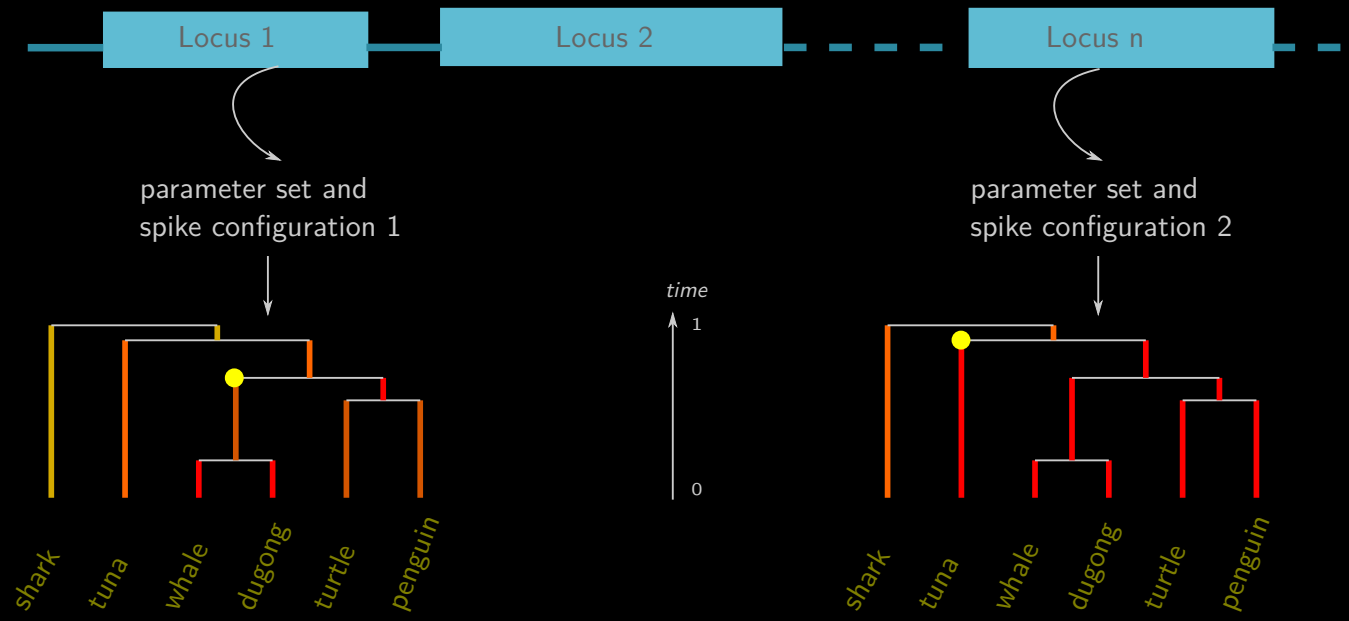
Looking specifically for ‘ecologically diverging genes’

Where we stand



- Inferences on simulations seem to work quite well.
- We may apply this to real data in the following months.

Hypothetical future progress for our understanding of evolution



- Two main objectives :

1. Date the tree with a relaxed clock.
2. Find loci showing a different pattern of spikes.

- Further hypothesis testing ideas :

1. Does divergence happen symmetrically at branching events ?
2. Does statistical signal support spikes happening at branching events ?

Take-home message

The strict clock hypothesis allowed the first tree dating.
Revealed slow (selectively stabilized) v.s. fast (neutral) loci

Relaxed clocks allow to spot acceleration/deceleration on the tree.
Revealed correlations of the clock with phenotypic traits.

Future clocks will hopefully be designed to bring further insights into the genomics of speciation.

Thank you for your attention !

- Key references :

Lartillot, N., Phillips, M. J., and Ronquist, F. (2016). A mixed relaxed clock model. *Phil. Trans. R. Soc. B*, 371(1699):20150132.

Lartillot, N. and Poujol, R. (2011). A phylogenetic model for investigating correlated evolution of substitution rates and continuous phenotypic characters. *Molecular biology and evolution*, 28(1):729–744.

Lepage, T., Bryant, D., Philippe, H., and Lartillot, N. (2007). A general comparison of relaxed molecular clock models. *Molecular biology and evolution*, 24(12):2669–2680.