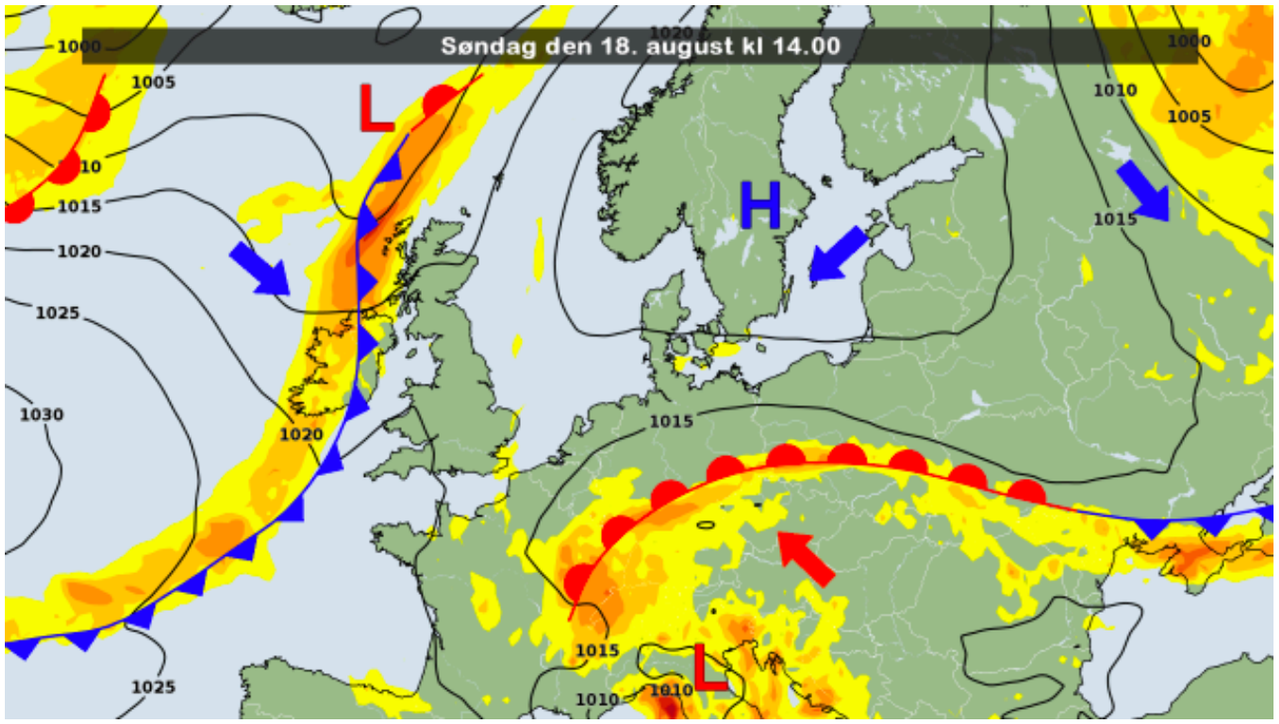
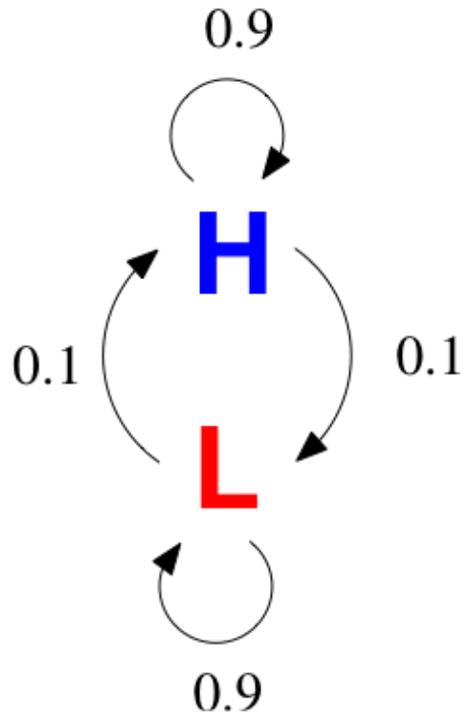
Hidden Markov models

Week 3

Kasper Munch

A weather model





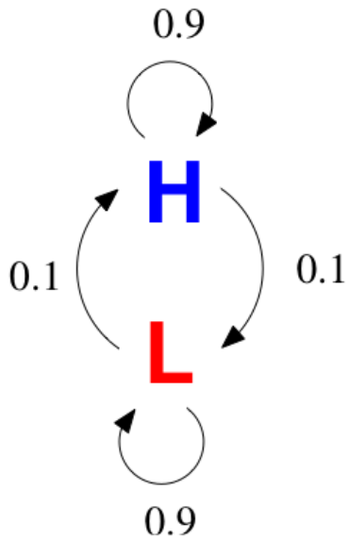
Mandag Tirsdag Onsdag Torsdag Fredag Lørdag Søndag Mandag Tirsdag Onsdag Torsdag Fredag

→ H → H → H → H → L → L → L → L → L → L → H → H →

Hidden Markov model

Transitions and emissions

Transition probabilities:

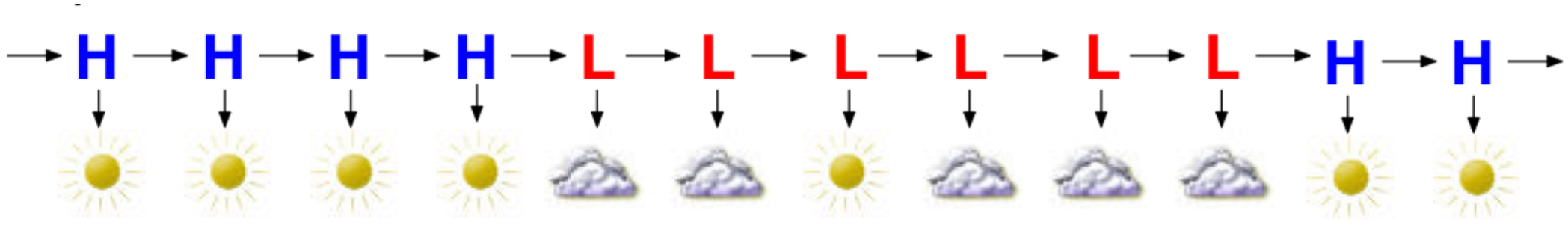


Emission probabilities:

H: $P(\text{☀}) = 0.7, \quad P(\text{☁}) = 0.3$

L: $P(\text{☀}) = 0.3, \quad P(\text{☁}) = 0.7$

Hidden Markov chain:

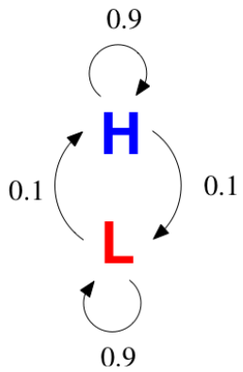


Watching the weather channel

Finding the most likely model parameters

Observations:

Model parameters:



$$\mathbf{H}: \quad P(\text{☀}) = 0.7, \quad P(\text{☁}) = 0.3$$

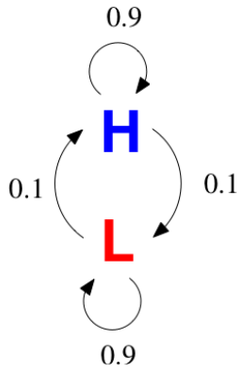
$$\mathbf{L}: \quad P(\text{☀}) = 0.3, \quad P(\text{☁}) = 0.7$$

Guessing air pressure yourself

Finding the most likely hidden states

Observations:

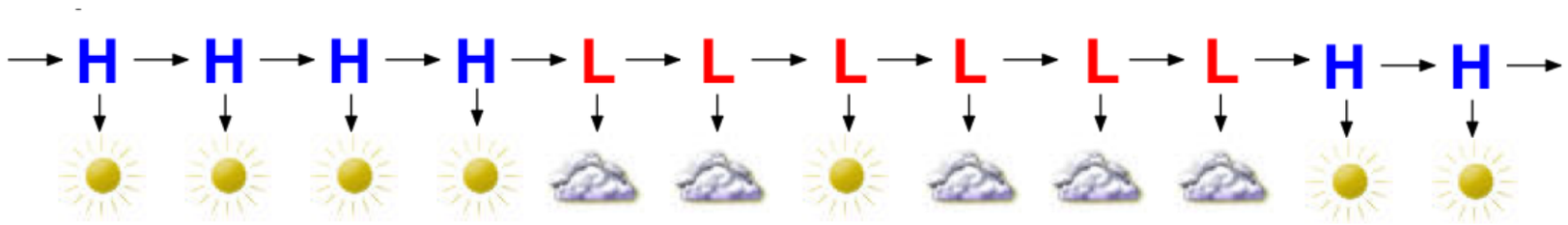
Model parameters:



$$\mathbf{H}: \quad P(\text{☀}) = 0.7, \quad P(\text{☁}) = 0.3$$

$$\mathbf{L}: \quad P(\text{☀}) = 0.3, \quad P(\text{☁}) = 0.7$$

Most likely sequence of hidden states:



Hidden Markov model

Why are they useful?

1. We can make the HMM **describe data** by optimising its **parameters**.

2. Using the HMM, we can estimate the **most likely sequence of hidden states**.

3. Using the HMM, we can estimate the **probability of each hidden state at each position** in the sequence

Hidden Markov model

Why are they useful?

2. We can make the HMM **describe data** by optimising its **parameters**.

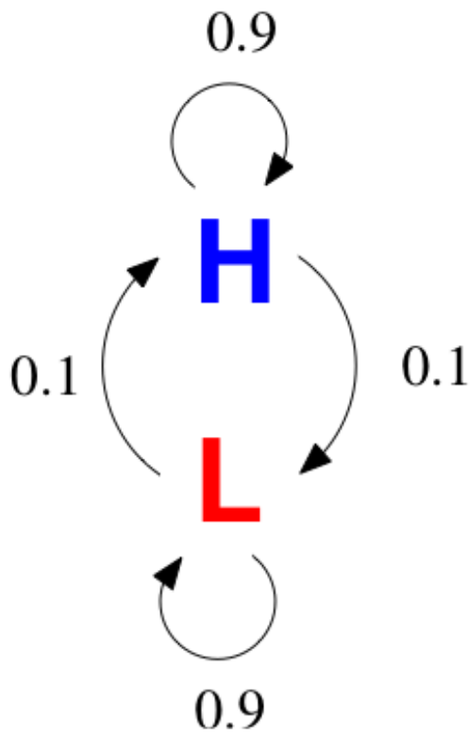
1. Using the HMM, we can estimate the **most likely sequence of hidden states**.

3. Using the HMM, we can estimate the **probability of each hidden state at each position** in the sequence

Markov chain

The Markov property:
The probability of the state at the next position in the chain only depends on the state at the current position.

I.e. the process has no memory.



Markov chain

$$\pi_1 \quad \pi_2 \quad \pi_3 \quad \pi_4 \quad \pi_5 \quad \pi_6 \quad \cdots \quad \pi_L$$

H → H → L → L → L → L → L → L

$$P(\pi) = P(\pi_L \,|\, \pi_{L-1})\; P(\pi_{L-1} \,|\, \pi_{L-2}) \;\; \cdots \;\; P(\pi_2 \,|\, \pi_1)\; P(\pi_1)$$

$$= P(\pi_1) \prod_{i=2}^{L} P(\pi_i \,|\, \pi_{i-1})$$

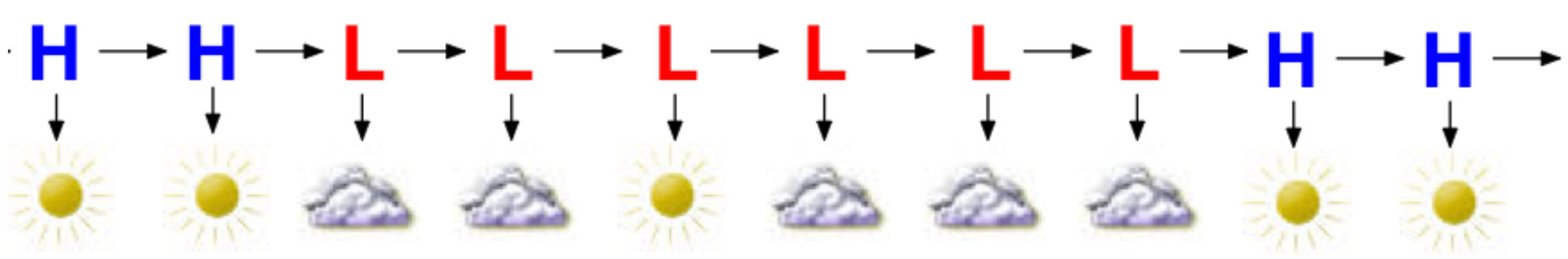
5 min discussion

- What is a state and what is an observable?
- What is hidden and what is observed?
- What is a Markov chain?

Hidden Markov chain

Hidden Markov chain: π

$\pi_1 \quad \pi_2 \quad \pi_3 \quad \pi_4 \quad \pi_5 \quad \pi_6 \quad \cdots$



$x_1 \quad x_2 \quad x_3 \quad x_4 \quad x_5 \quad x_6 \quad \cdots$

Observed sequence: x

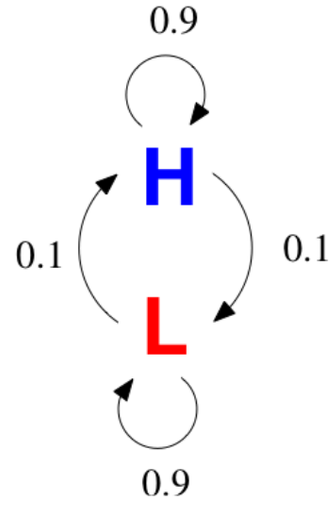
Hidden Markov model

State path: $\pi_1 \quad \pi_2 \quad \pi_3 \quad \pi_4 \quad \pi_5 \quad \pi_6 \quad \cdots$

Sequence symbols: $x_1 \quad x_2 \quad x_3 \quad x_4 \quad x_5 \quad x_6 \quad \cdots$

Transition probability:

$$a_{kl} = P(\pi_i = l \,|\, \pi_{i-1} = k)$$



Emission probability:

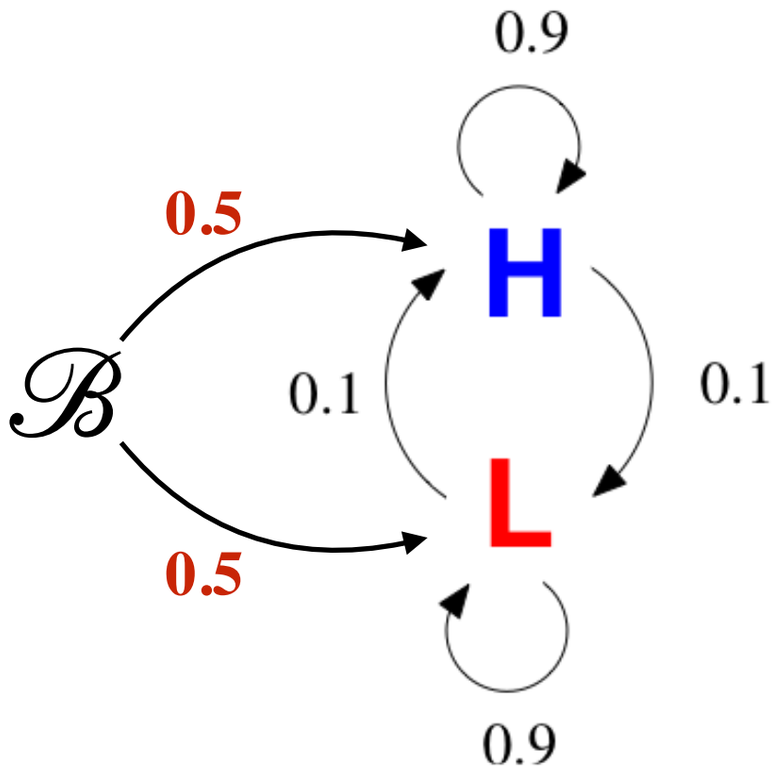
$$e_k(b) = P(x_i = b \,|\, \pi_i = k)$$

H: $P(\text{☀}) = 0.7, \quad P(\text{☁}) = 0.3$

L: $P(\text{☀}) = 0.3, \quad P(\text{☁}) = 0.7$

A begin state

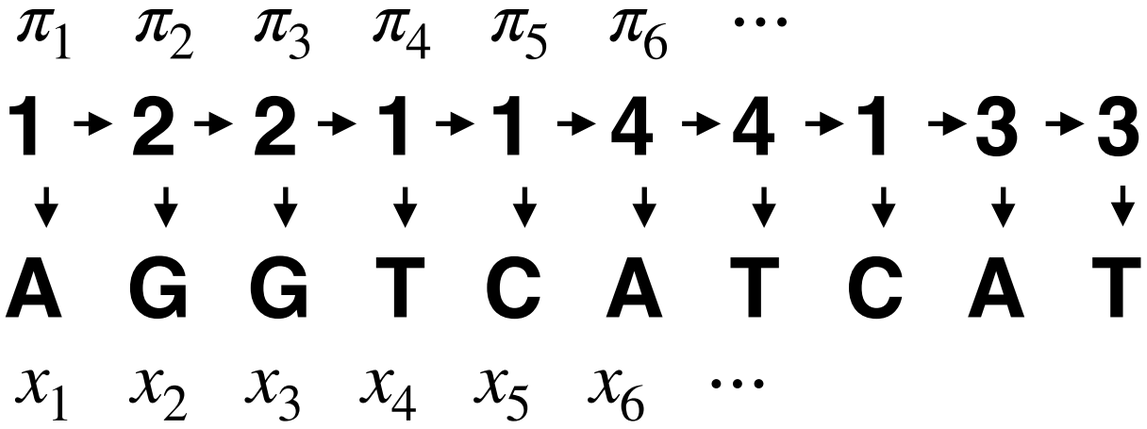
Transition probabilities from begin state reflect prior belief:

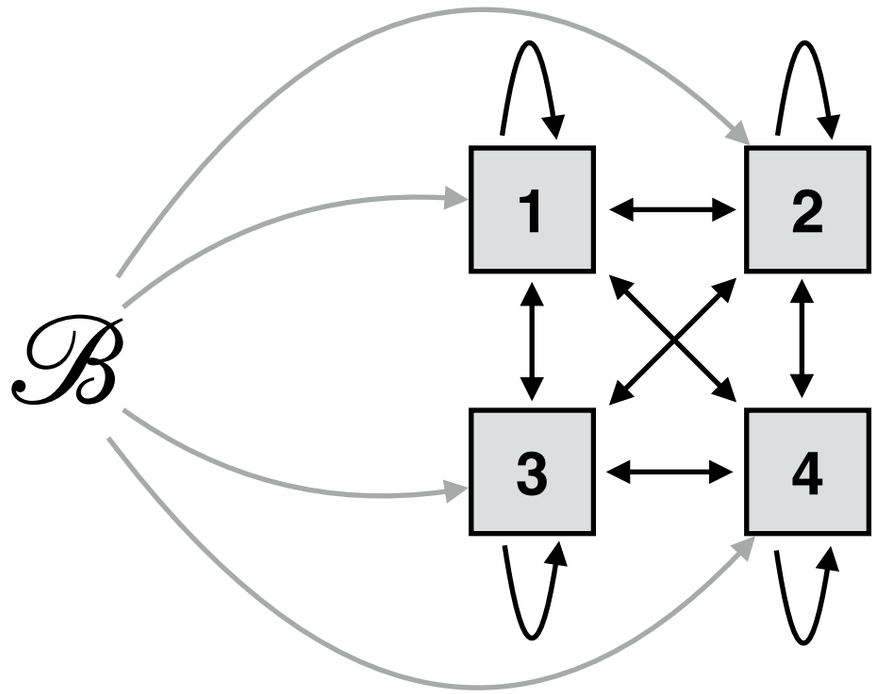


Transition probabilities from the null state represents prior belief

Joint probability

Of the observables x and hidden states π



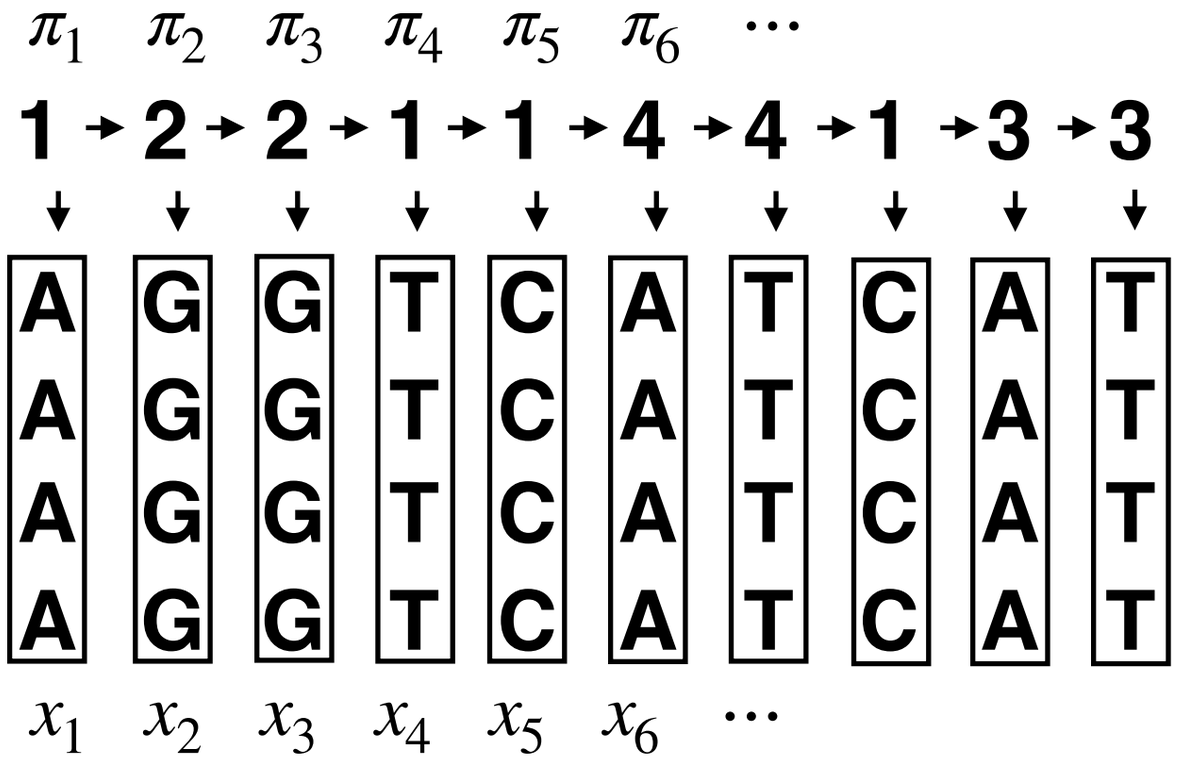


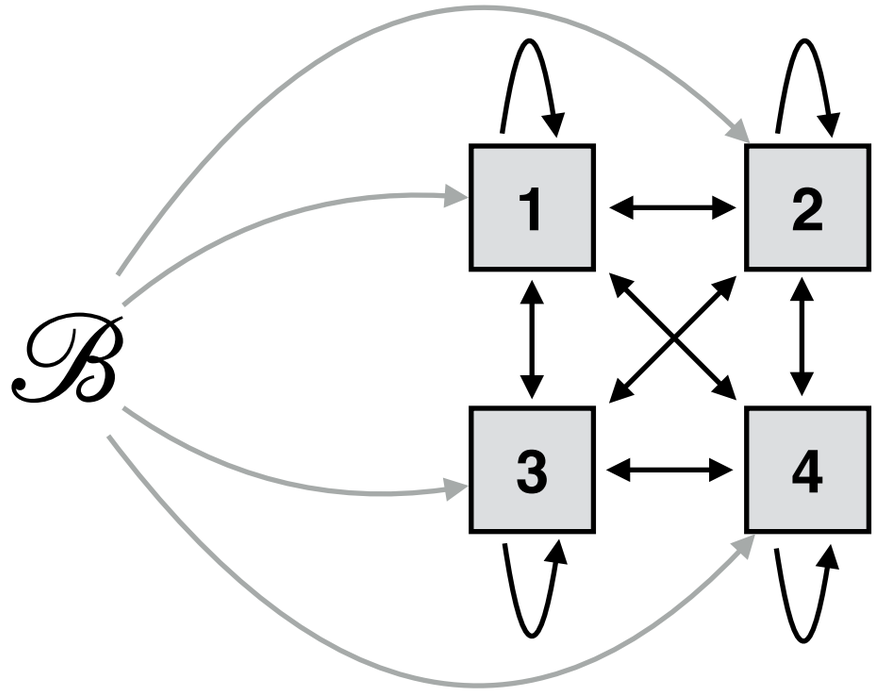
Joint probability of x and π:

$$P(x, \pi) = a_{0\pi_1} \prod_{i=1}^{L} e_{\pi_i}(x_i) a_{\pi_i \pi_{i+1}}$$

Joint probability

Of the observables x and hidden states π





$$P(x, \pi) = a_{0\pi_1} \prod_{i=1}^{L} e_{\pi_i}(x_i) a_{\pi_i \pi_{i+1}}$$

The probability of our sequence

Across *all* possible hidden state paths

How likely is it that our observables are generated by our hidden Markov model?

The probability of our sequence

Across *all possible* hidden state paths

$$P(x) = \sum_{\pi} P(x, \pi) \qquad P(x, \pi) = a_{0\pi_1} \prod_{i=1}^{L} e_{\pi_i}(x_i) a_{\pi_i \pi_{i+1}}$$

The probability of our sequence

Across *all possible* hidden state paths

$$P(x) = \sum_{\pi} P(x, \pi) \qquad P(x, \pi) = a_{0\pi_1} \prod_{i=1}^{L} e_{\pi_i}(x_i) a_{\pi_i \pi_{i+1}}$$

Problem: the number of paths grows exponentially with sequence length!

Luckily, we can use dynamic programming like in the Viterbi algorithm.

The probability of our sequence

Across *all possible* hidden state paths

$f_k(i)$: The joint **probability** of **all state paths** for the first i states and i observables, assuming that the last state is k.

The probability of our sequence

Across *all possible* hidden state paths

$f_k(i)$: The joint **probability** of **all state paths** for the first i states and i observables, assuming that the last state is k.

Joint probability of all paths up to i: $f_l(i)$

$$\begin{array}{cccccc} \pi_1 & \pi_2 & \pi_3 & \pi_4 & \pi_5 & \cdots \\ x_1 & x_2 & x_3 & x_4 & x_5 & \cdots \end{array} \qquad \begin{array}{c} \pi_i = l \\ x_i \end{array}$$

Most probable state path

Because of the Markov property we can compute its probability recursively:

$$f_l(i) = \sum_k f_k(i-1)\, a_{kl}\, e_l(x_i)$$

Joint probability of optimal path up to $i-1$: $f_k(i-1)$

Transition: a_{kl}

Emission: $e_l(x_i)$

π_1	π_2	π_3	π_4	π_5	$\cdots$	$\pi_{i-1} = k$	$\pi_i = l$
x_1	x_2	x_3	x_4	x_5	$\cdots$	x_{i-1}	x_i

Joint probability of optimal path up to $i-1$: $f_1(i-1)$

Transition: a_{1l}

Emission: $e_l(x_i)$

π_1	π_2	π_3	π_4	π_5	$\cdots$	$\pi_{i-1} = 1$	$\pi_i = l$
x_1	x_2	x_3	x_4	x_5	$\cdots$	x_{i-1}	x_i

Joint probability of optimal path up to $i-1$: $f_2(i-1)$

Transition: a_{1l}

Emission: $e_l(x_i)$

π_1	π_2	π_3	π_4	π_5	$\cdots$	$\pi_{i-1} = 1$	$\pi_i = l$
x_1	x_2	x_3	x_4	x_5	$\cdots$	x_{i-1}	x_i

Joint probability of optimal path up to $i-1$: $f_3(i-1)$

Transition: a_{1l}

Emission: $e_l(x_i)$

π_1	π_2	π_3	π_4	π_5	$\cdots$	$\pi_{i-1} = 1$	$\pi_i = l$
x_1	x_2	x_3	x_4	x_5	$\cdots$	x_{i-1}	x_i

Joint probability of optimal path up to $i-1$: $f_4(i-1)$

Transition: a_{1l}

Emission: $e_l(x_i)$

π_1	π_2	π_3	π_4	π_5	$\cdots$	$\pi_{i-1} = 1$	$\pi_i = l$
x_1	x_2	x_3	x_4	x_5	$\cdots$	x_{i-1}	x_i

SUM over $k \in \{1, 2, 3, 4\}$ at position $i-1$

$$f_l(i)$$

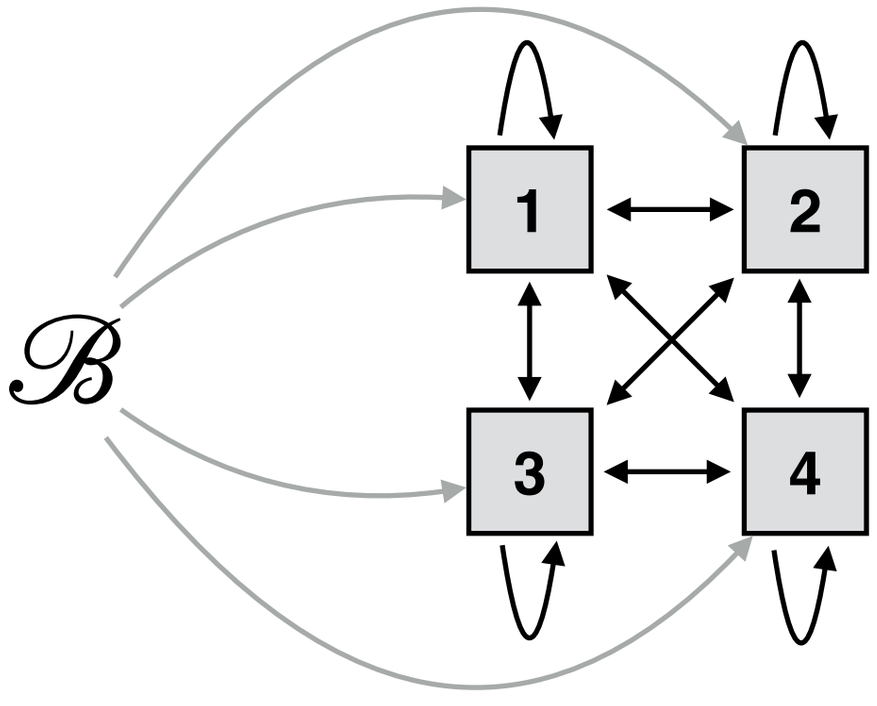
Recursion

Recursive case:

$$f_l(i) = e_l(x_i) \sum_k (f_k(i-1)\, a_{kl})$$

Base case: when no sequence is used the state is *always* the null state:

$$v_0(\mathscr{B}) = 1$$



Recursion

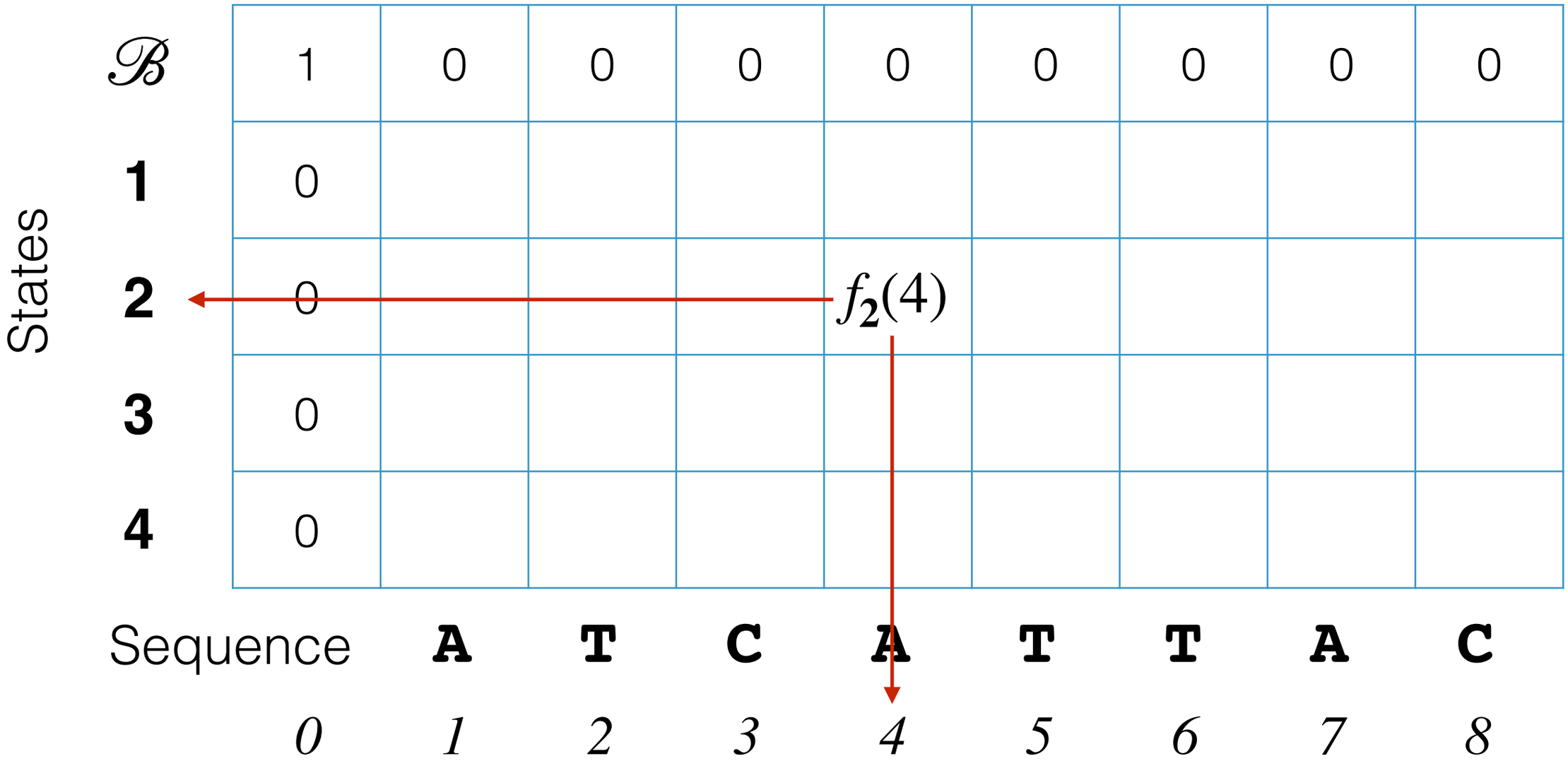
$$f_l(i) = \sum_k f_k(i-1)\, a_{kl}\, e_l(x_i)$$

How do we avoid computing the same probability many many times?

Remember pairwise alignment?

Dynamic programming table

for states and observables



Dynamic programming table

for states and observables

$\mathscr{B}$		1	0	0	0	0	0	0	0	0
States	**1**	0								$f_1(8)$
	2	0								$f_2(8)$
	3	0								$f_3(8)$
	4	0								$f_4(8)$
Sequence			**A**	**T**	**C**	**A**	**T**	**T**	**A**	**C**

Sum

Dynamic programming table

for states and observables

States	$\mathscr{B}$	1	0	0	0	0	0	0	0	0
	1	0			$f_1(3)$					
	2	0			$f_2(3)$	$f_2(4)$				
	3	0			$f_3(3)$					
	4	0			$f_4(3)$					
Sequence		**A**	**T**	**C**	**A**	**T**	**T**	**A**	**C**	

$f_2(4)$ is the sum of joint probabilities for all hidden state paths ending at state 2 at poisition 3.

Joint probability of optimal path up to $i-1$: $f_k(i-1)$

Transition: a_{kl}

Emission: $e_l(x_i)$

π_1	π_2	π_3	π_4	π_5	$\cdots$	$\pi_{i-1} = k$	$\pi_i = l$
x_1	x_2	x_3	x_4	x_5	$\cdots$	x_{i-1}	x_i

Dynamic programming table

for states and observables

States

$\mathscr{B}$	1	0	0	$f_{\mathscr{B}}(3)$	0	0	0	⌒
1	0			$f_1(3)$				
2	0			$f_2(3)$	$v_2(4)$			
3	0			$f_3(3)$				
4	0			$f_4(3)$				

Sequence **A T C A T T A C**

Formally, we sum over all k including $\mathscr{B}$, but since $f_{\mathscr{B}}(i) = 0$ for $i > 0$, we can safely ignore it.

Joint probability of optimal path up to $i-1$: $f_k(i-1)$

Transition: a_{kl}

Emission: $e_l(x_i)$

$$\begin{array}{cccccc} \pi_1 & \pi_2 & \pi_3 & \pi_4 & \pi_5 & \cdots \\ x_1 & x_2 & x_3 & x_4 & x_5 & \cdots \end{array} \quad \begin{array}{cc} \pi_{i-1} = k & \pi_i = l \\ x_{i-1} & x_i \end{array}$$

Dynamic programming table

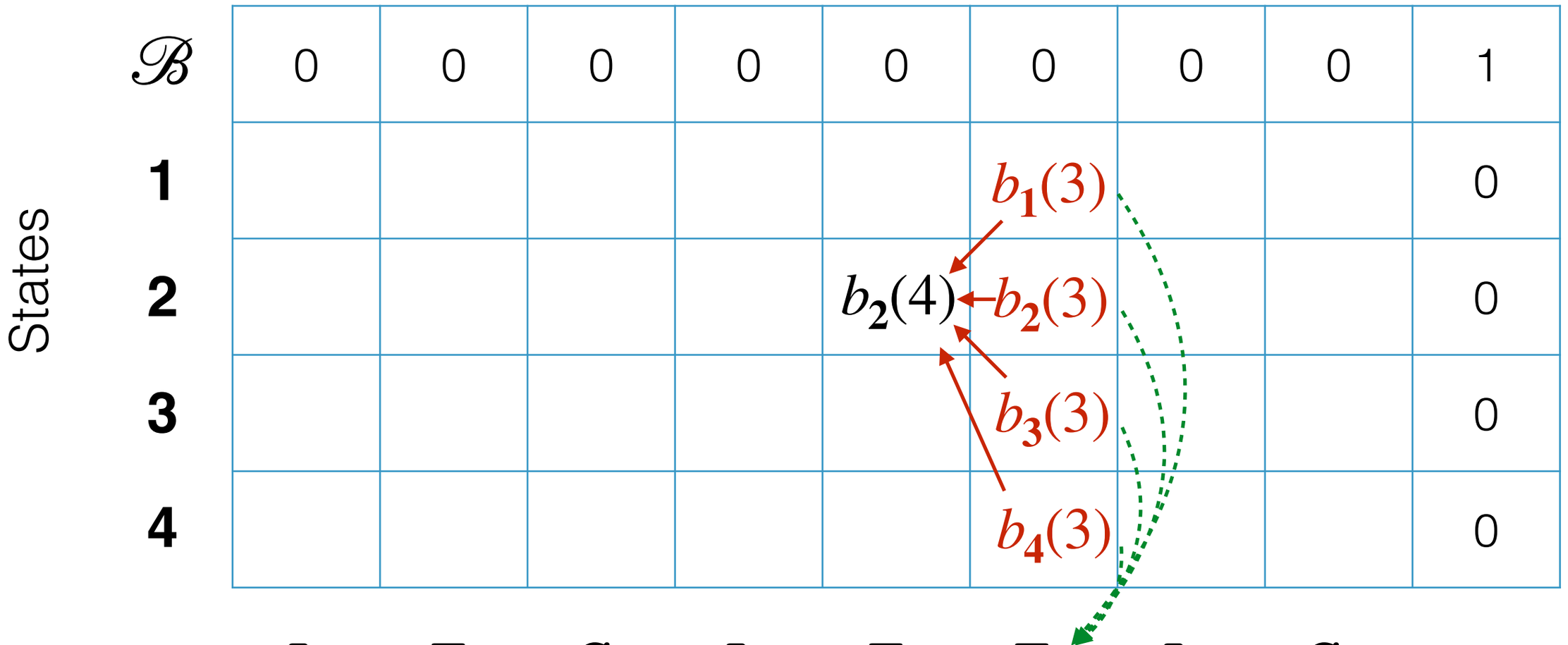
for states and observables

$\mathscr{B}$		1	0	0	0	0	0	0	0	0
States	**1**	0								$f_1(8)$
	2	0								$f_2(8)$
	3	0								$f_3(8)$
	4	0								$f_4(8)$
	Sequence		**A**	**T**	**C**	**A**	**T**	**T**	**A**	**C**

Fill the table column by column, from left to right

Sum

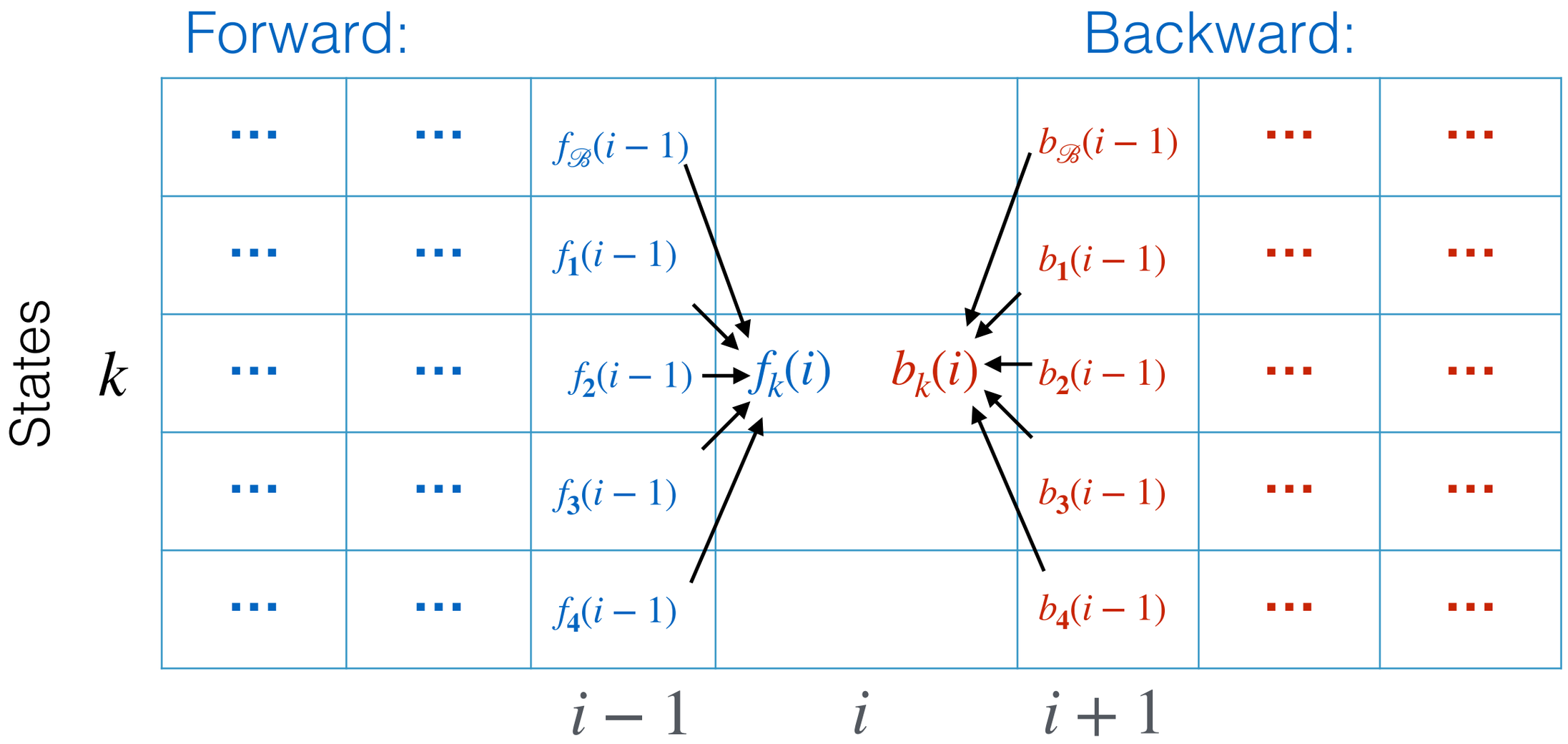
Backward algorithm



$$b_l(i) = \sum_k a_{kl}\, e_l(x_{i+1})\, b_k(i+1)$$

Posterior probability

- of a state at a position in the sequence



The proportion of the total probability contributed by paths with state k at position i:

$$P(\pi_i = k \,|\, x) = \frac{P(\pi_i = k)}{P(x)} = \frac{f_k(i)\, b_k(i)}{P(x)}$$

Posterior decoding

Hidden states with max posterior probability

Posterior probability that the state at position i is k:

$$P(\pi_i = k \,|\, x) = \frac{f_k(i)\, b_k(i)}{P(x)}$$

Important: Contributions from many different paths. So the sequence of states with the highest posterior probability will not correspond to a valid path through the state model.

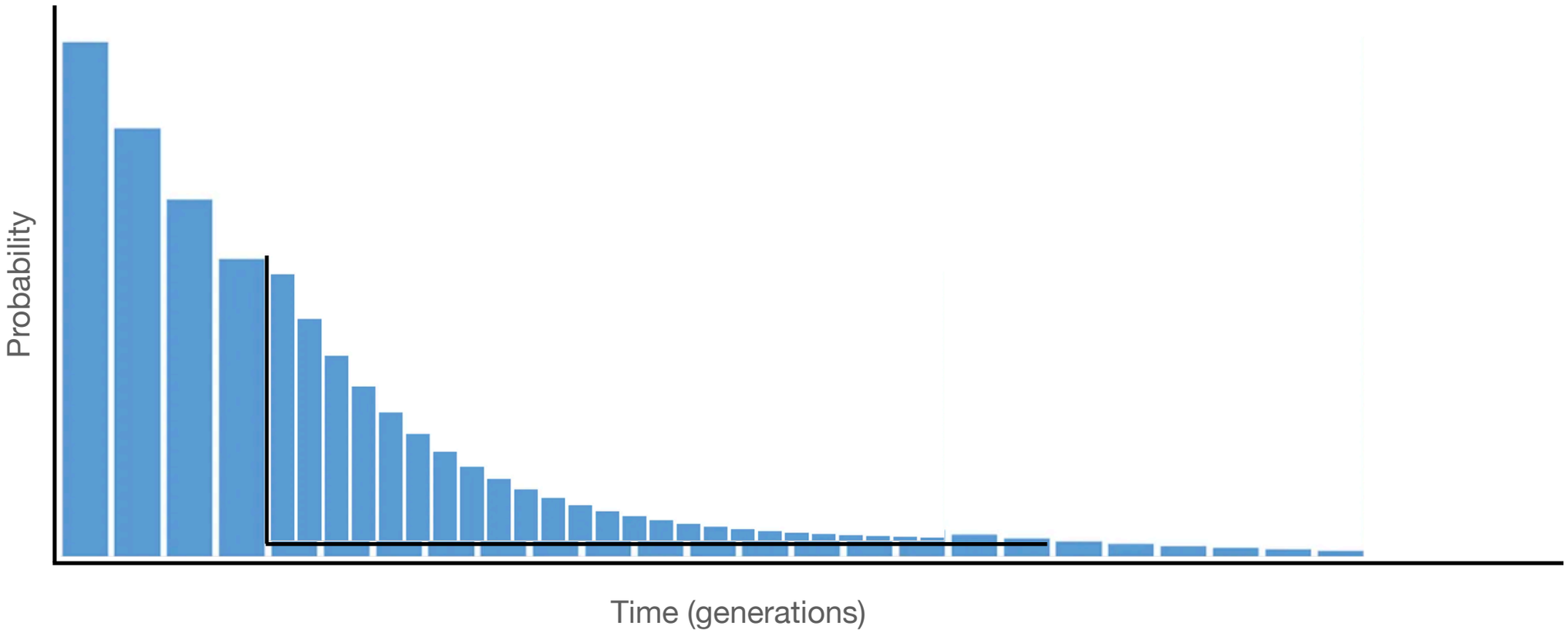
Demography with PSMC

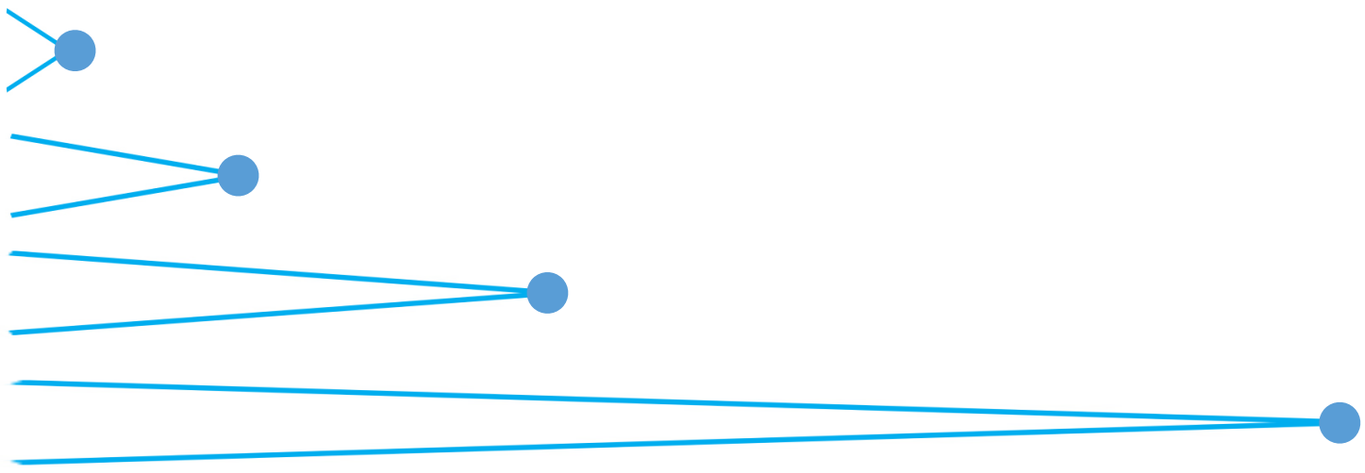
Week 3

Kasper Munch

Distribution of TMRCA

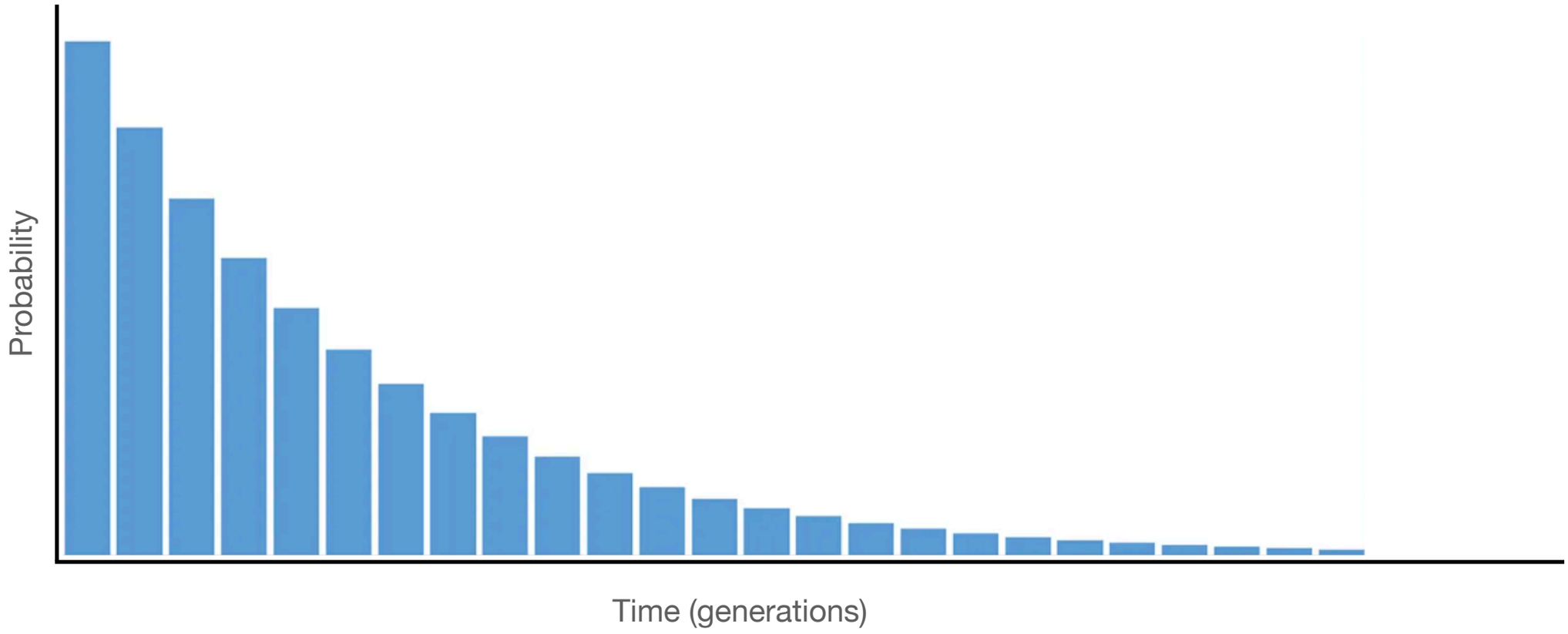
For two sequences

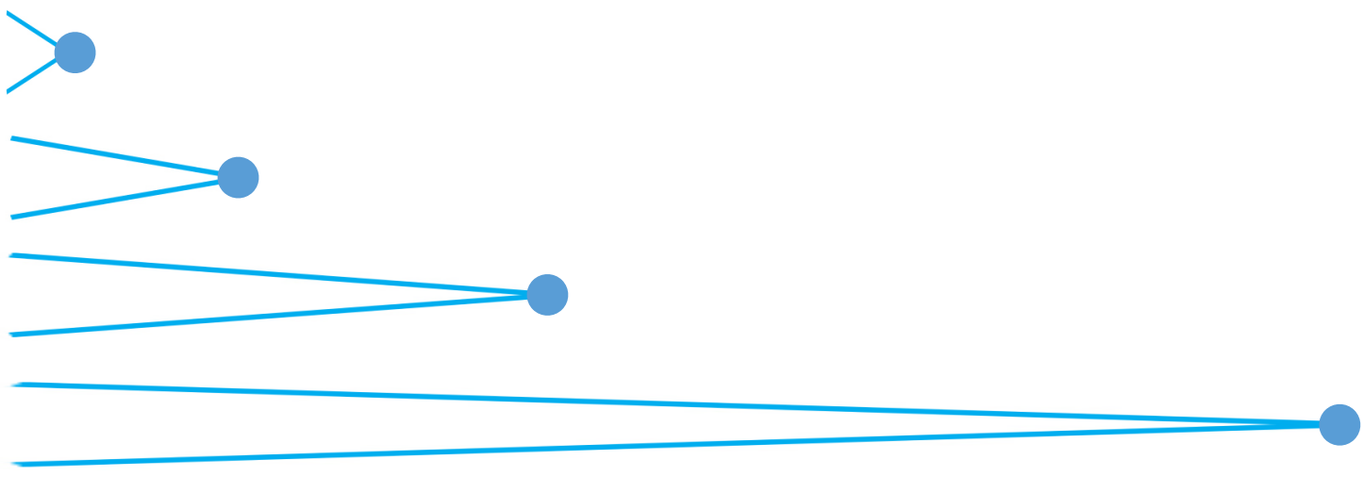




Distribution of TMRCA

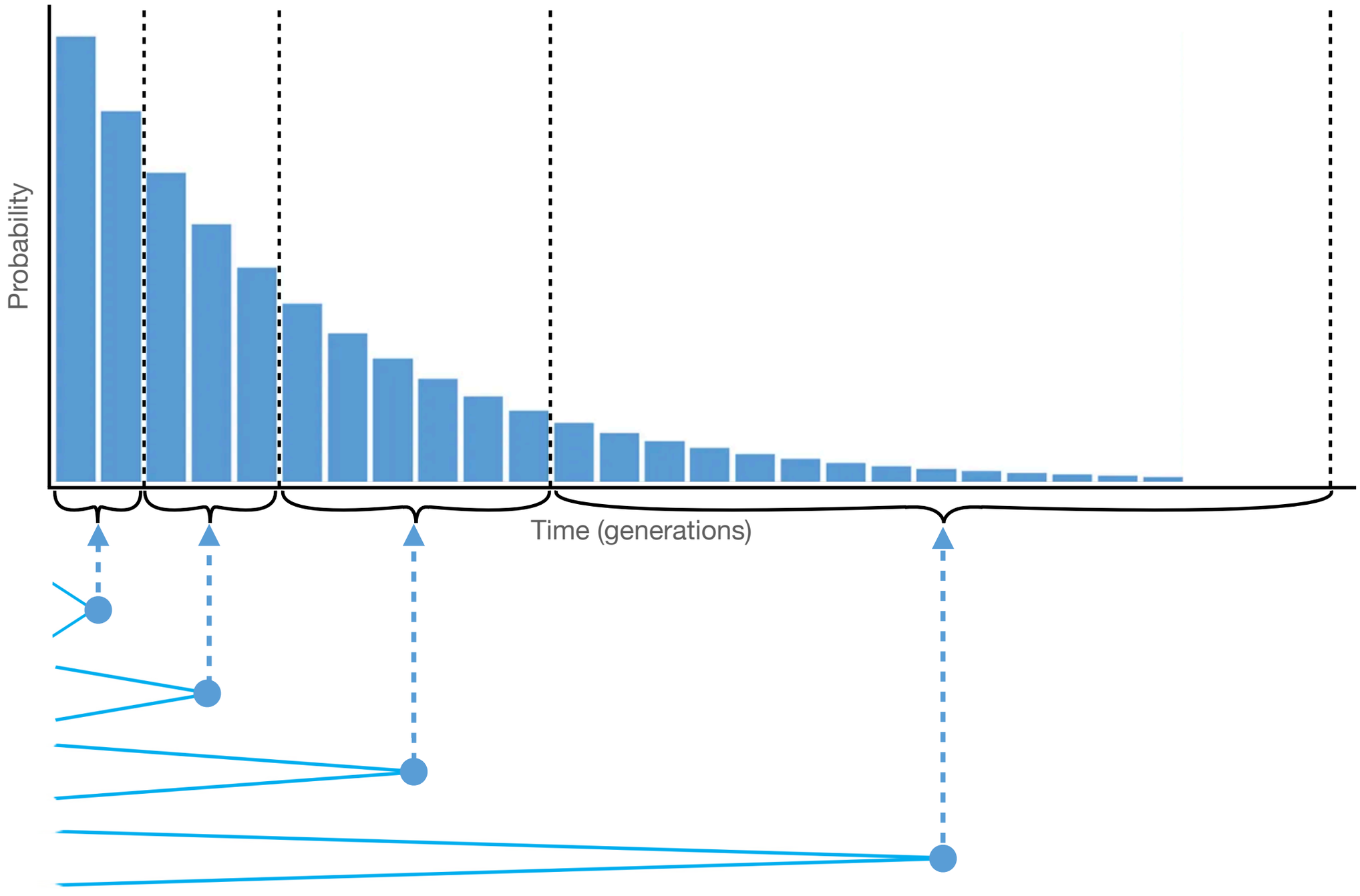
For two sequences





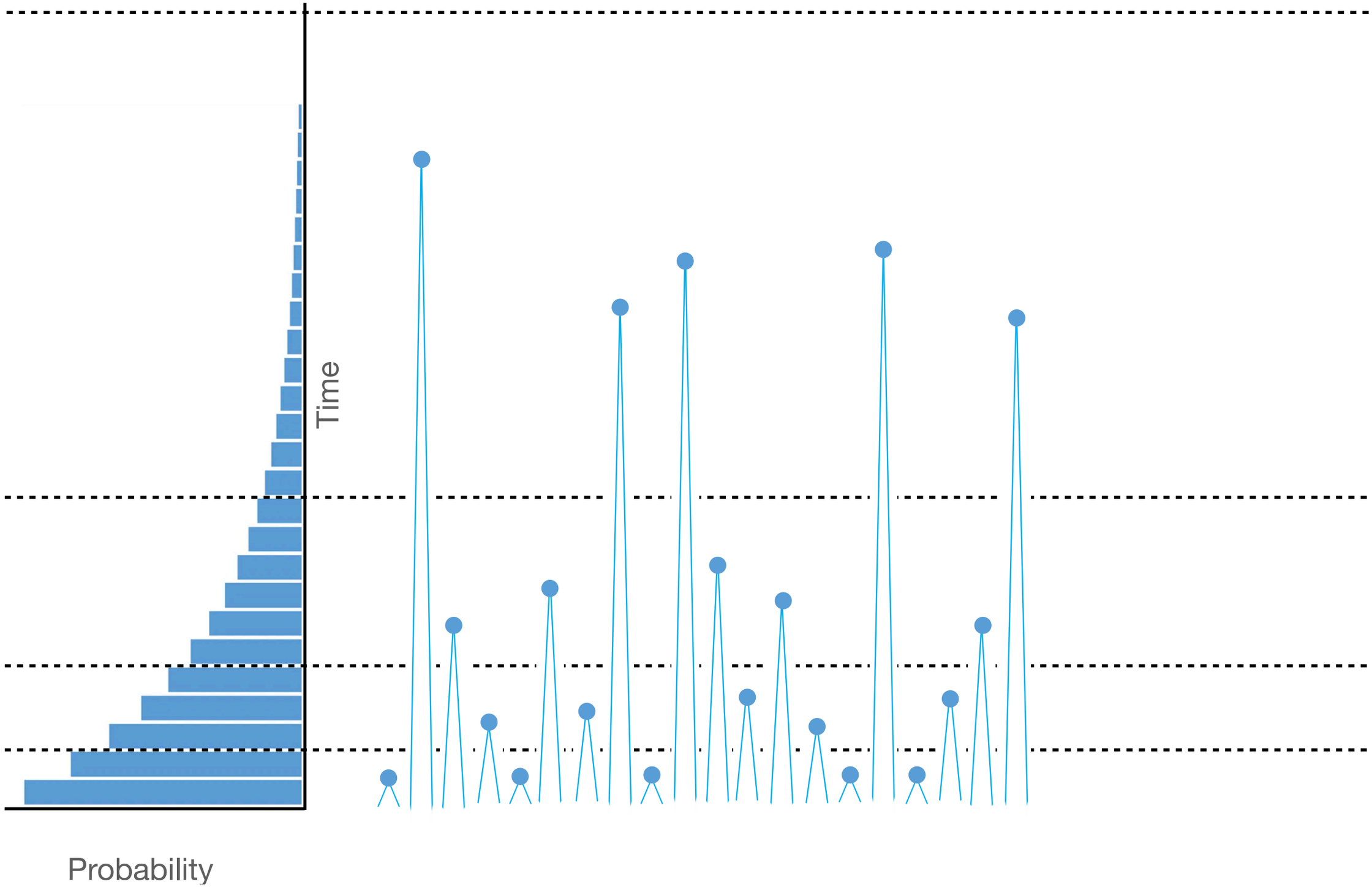
Logarithmic bins

Makes coalescence in each bin equally likely



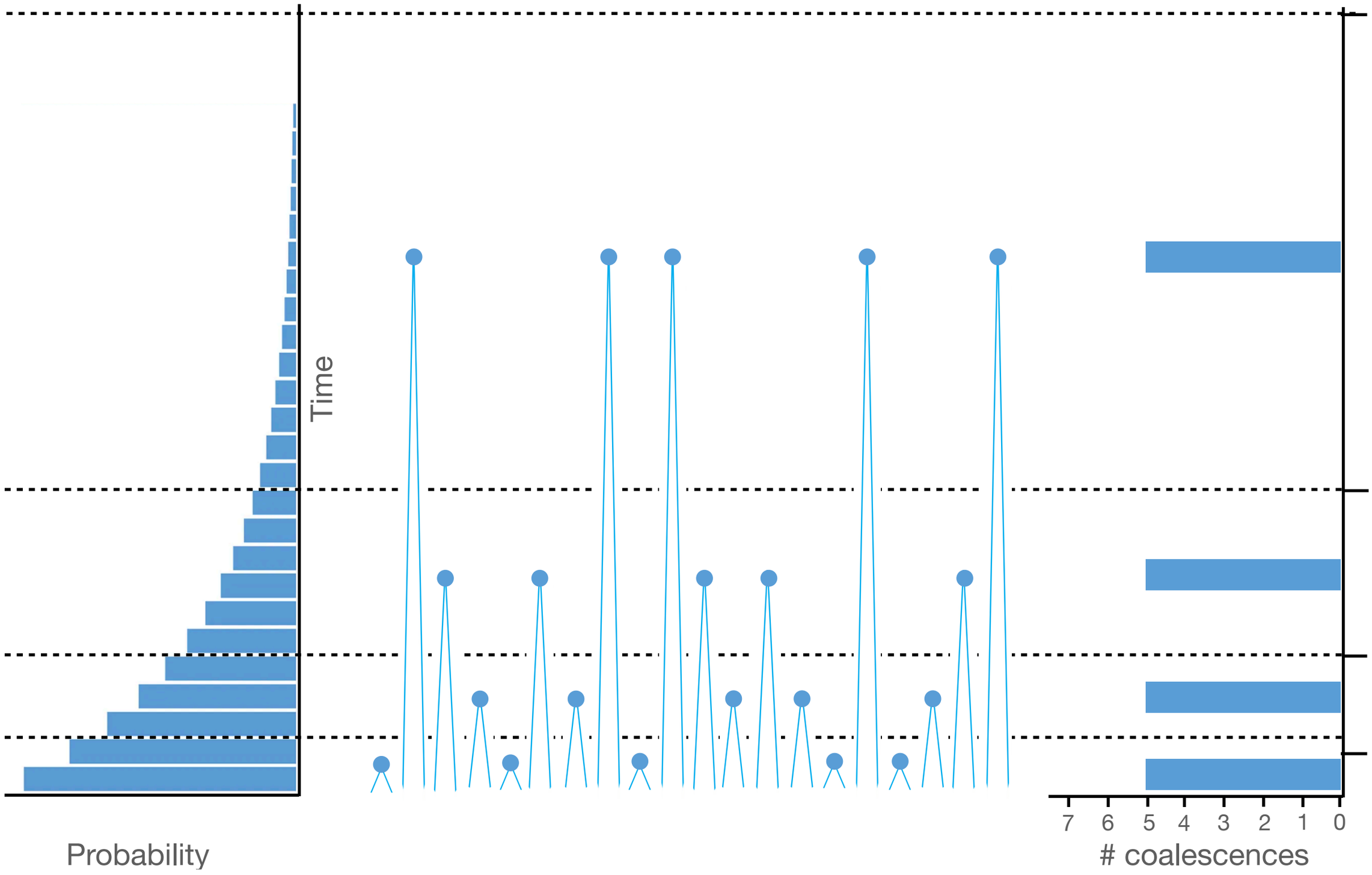
Thought experiment

Suppose we knew the TMRCA of each sequence segment



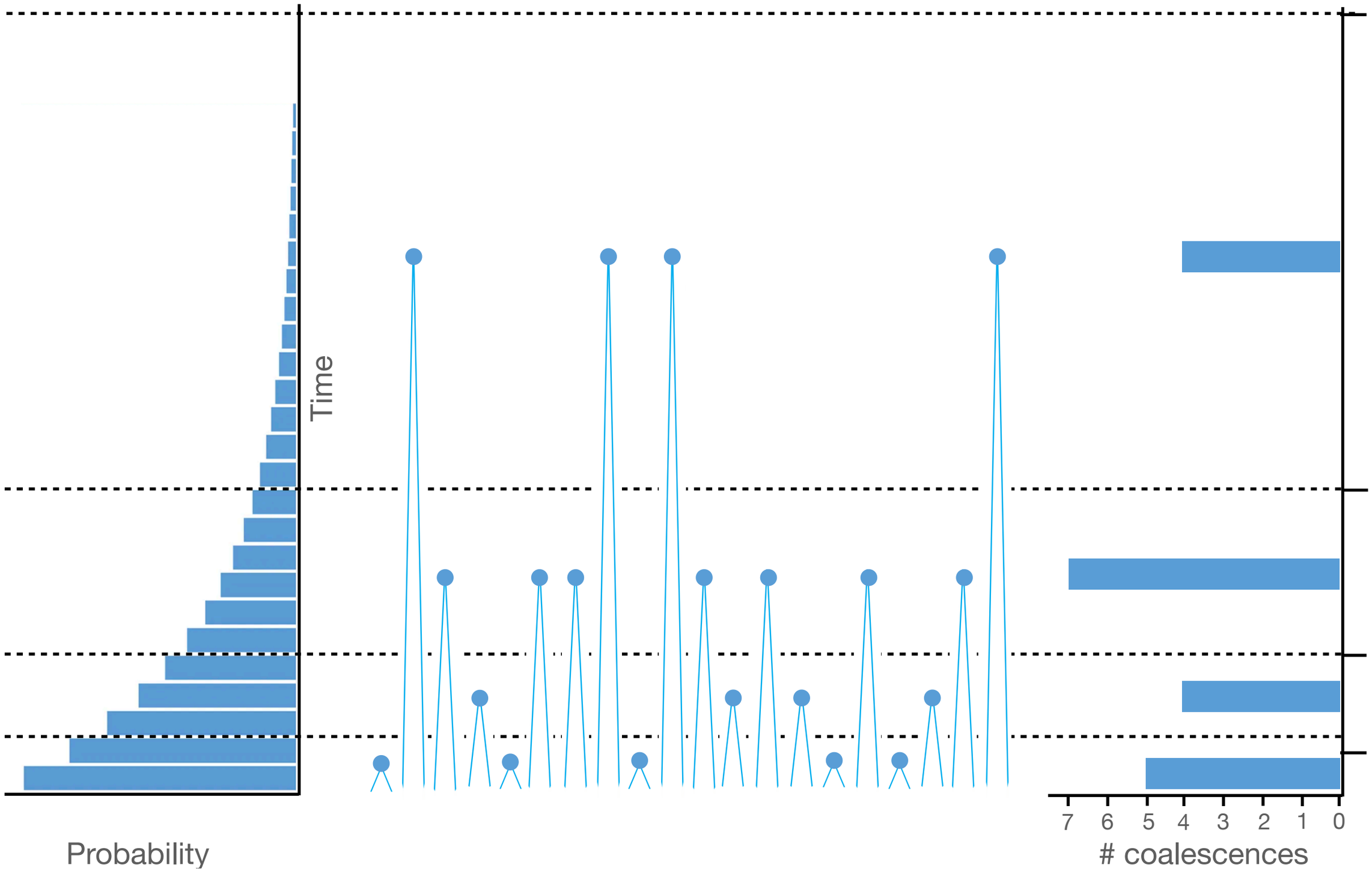
Thought experiment

Suppose we knew the TMRCA of each sequence segment



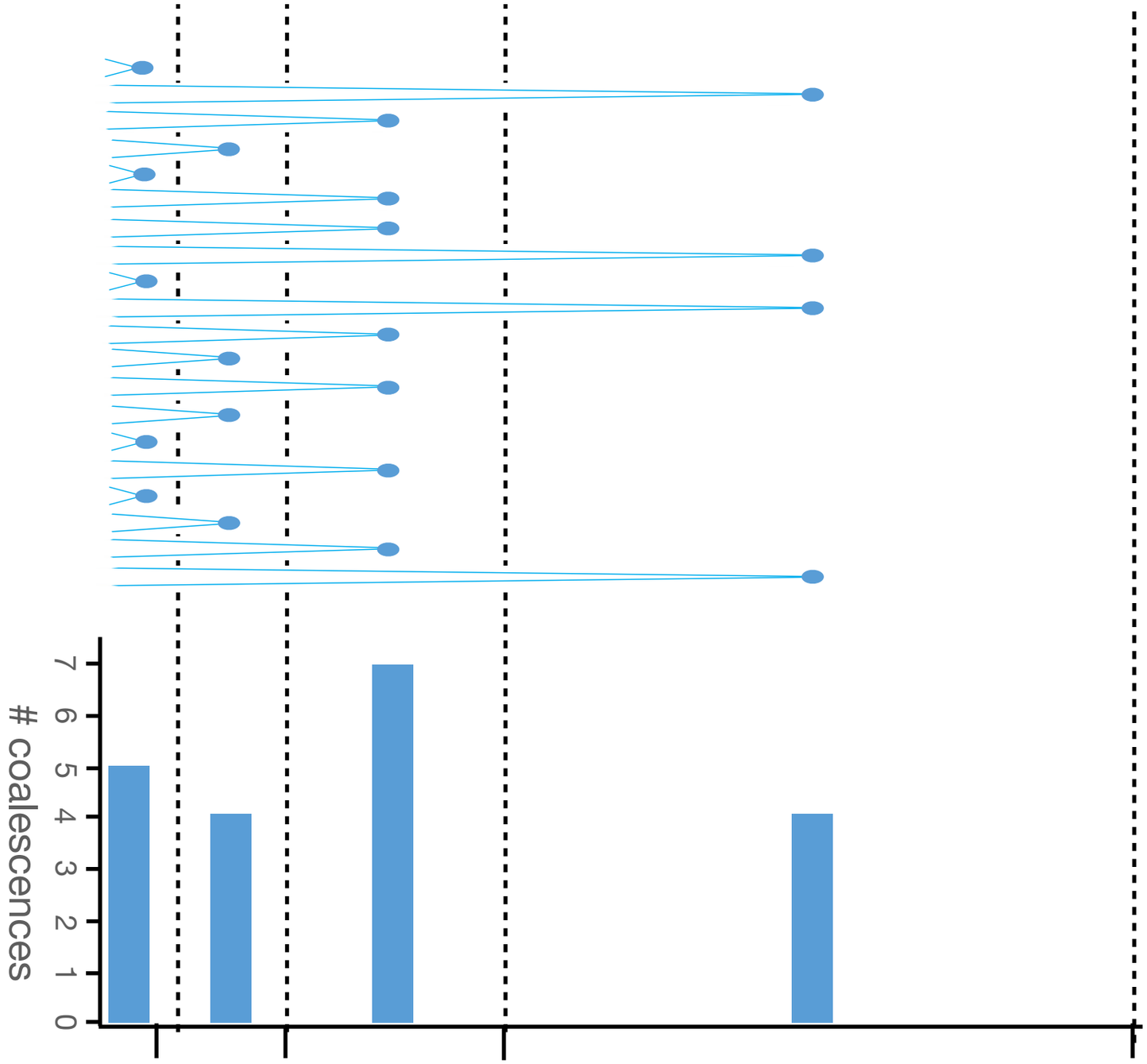
Thought experiment

Suppose we knew the TMRCA of each sequence segment



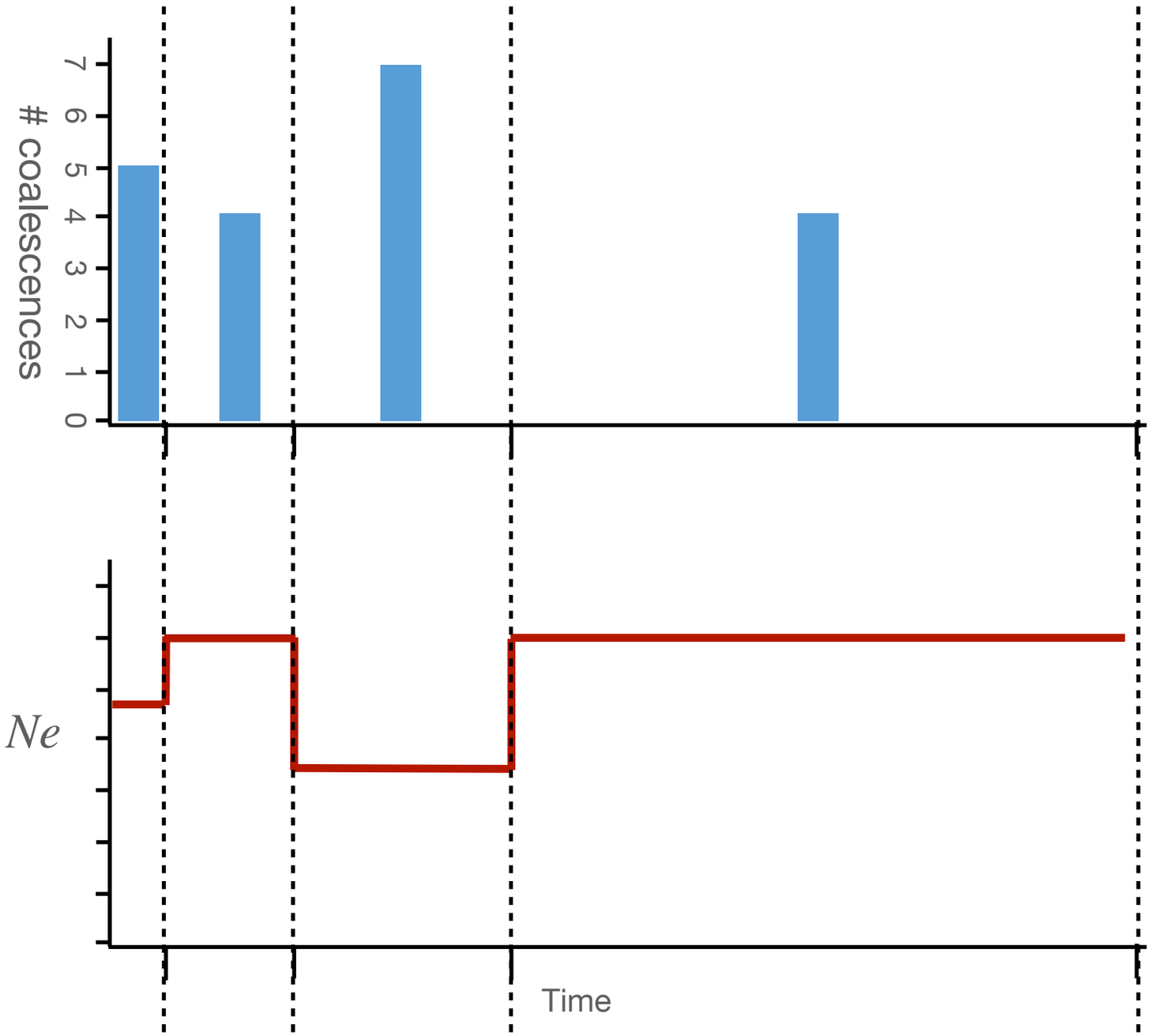
Thought experiment

Suppose we knew the TMRCA of each sequence segment



Thought experiment

Suppose we knew the TMRCA of each sequence segment



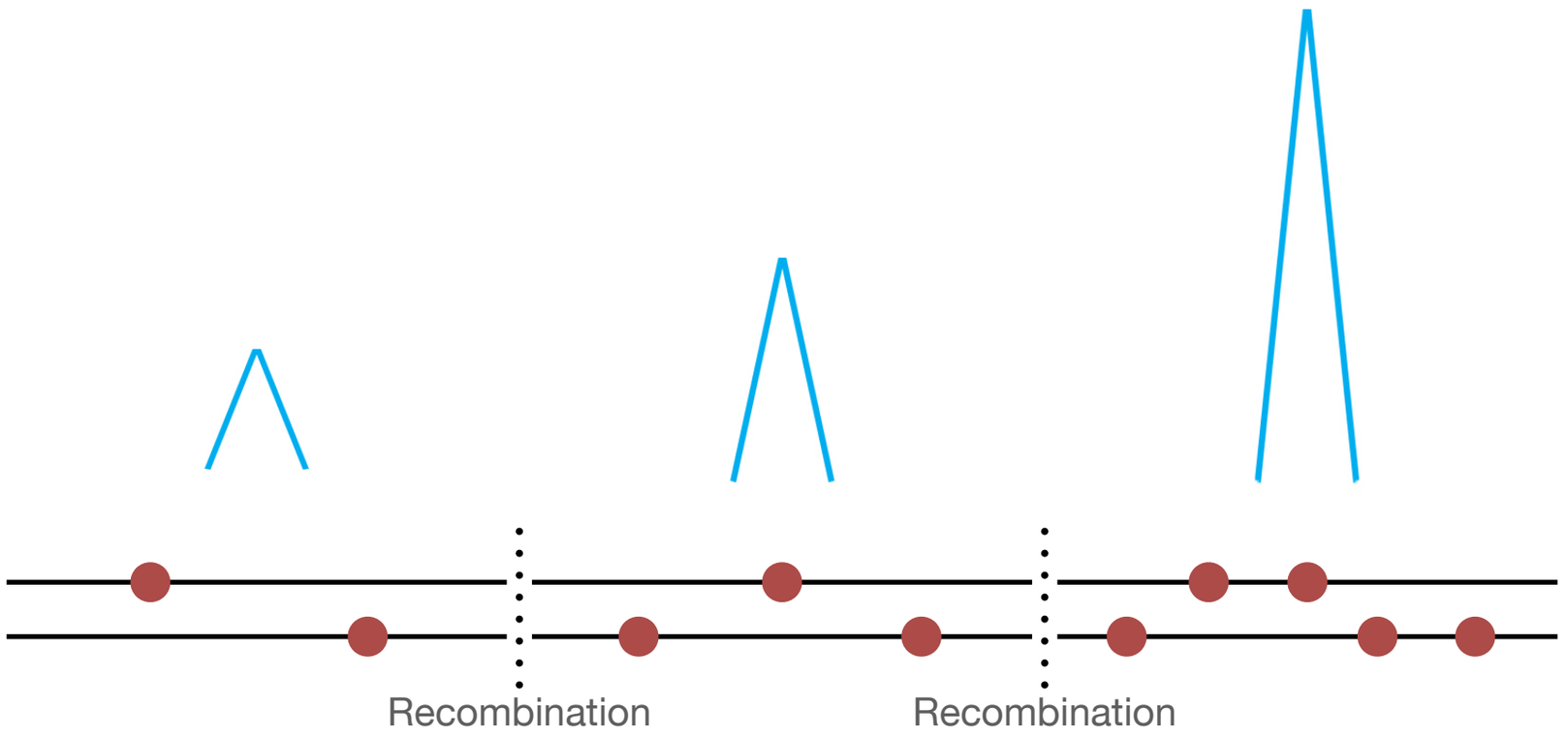
Back to reality

- But don’t despair

- We **do not** have:
  - The segments between recombination events.
  - The TMRCA of each segment.
- We **do** have:
  - The sequence differences.
  - A model (the ARG) that relate sequence differences to segment length and TMRCA.

Mutations and TMRCA

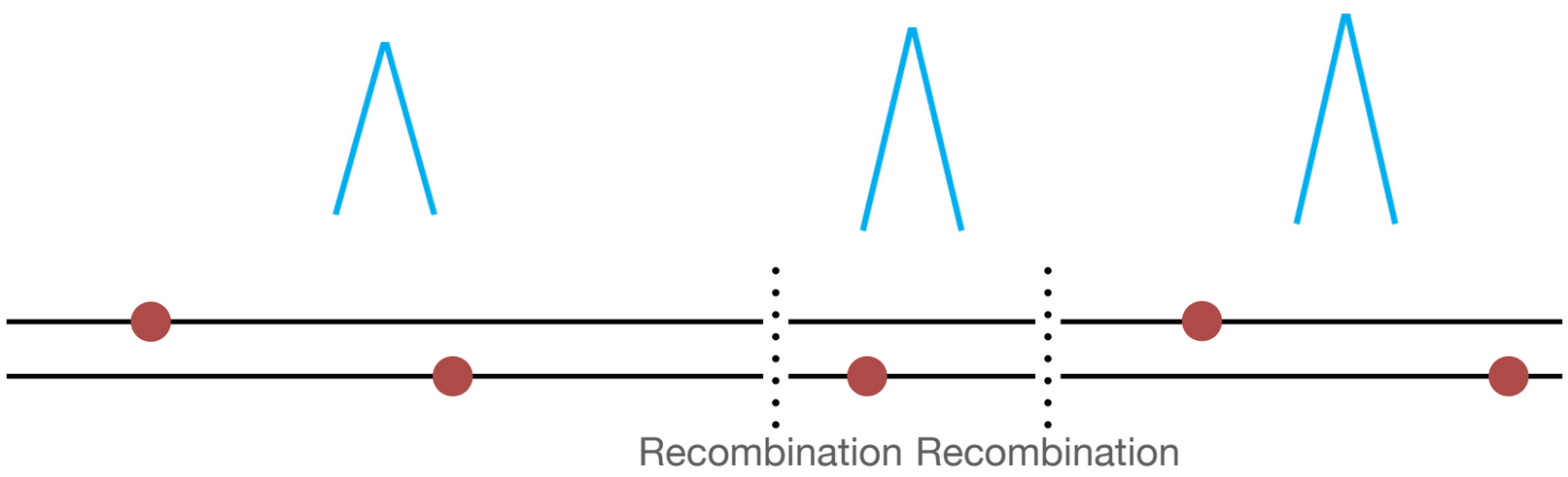
Which segment has the longest TMRCA?



- Why?

Mutations and TMRCA

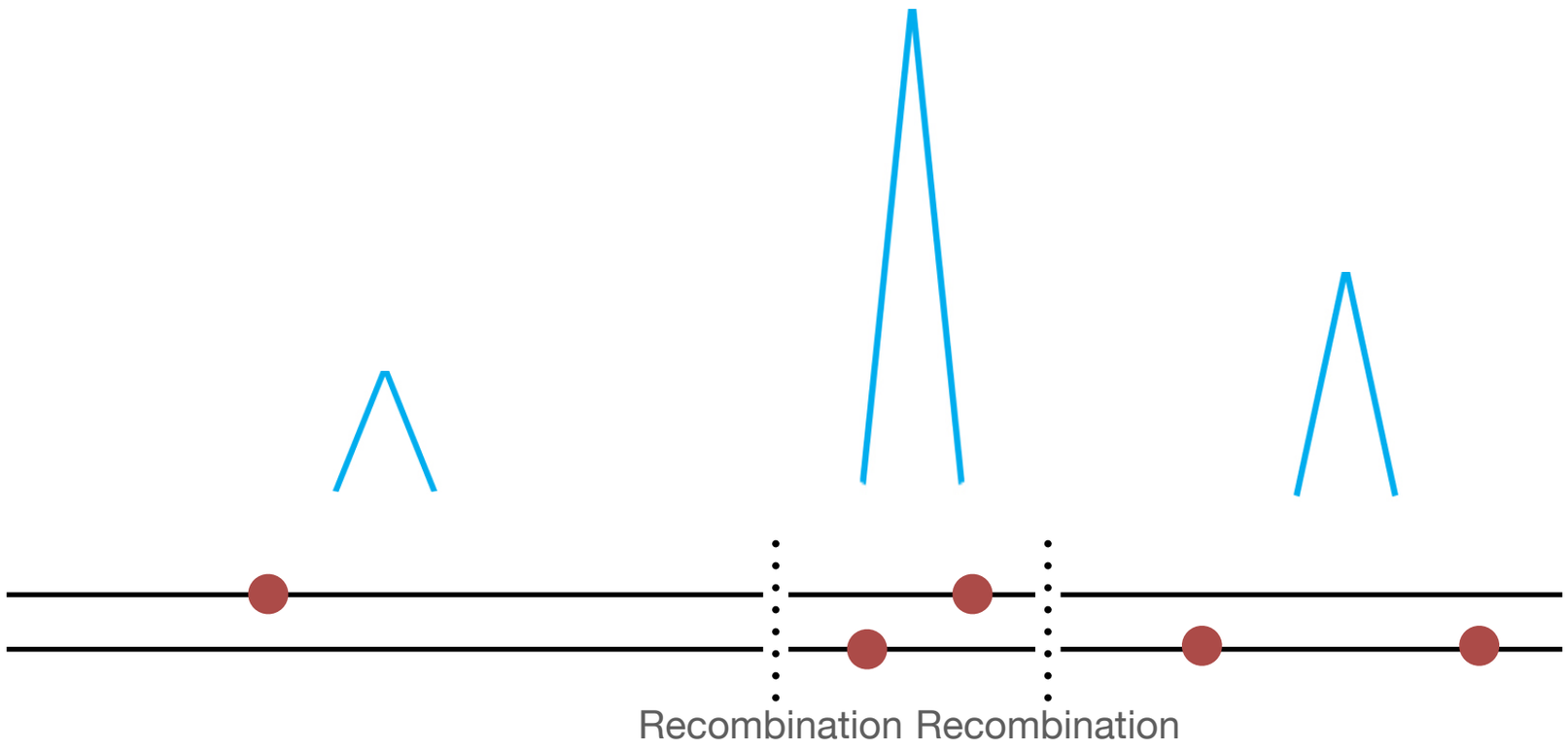
Which segment has the longest TMRCA?



- Why?

Mutations and TMRCA

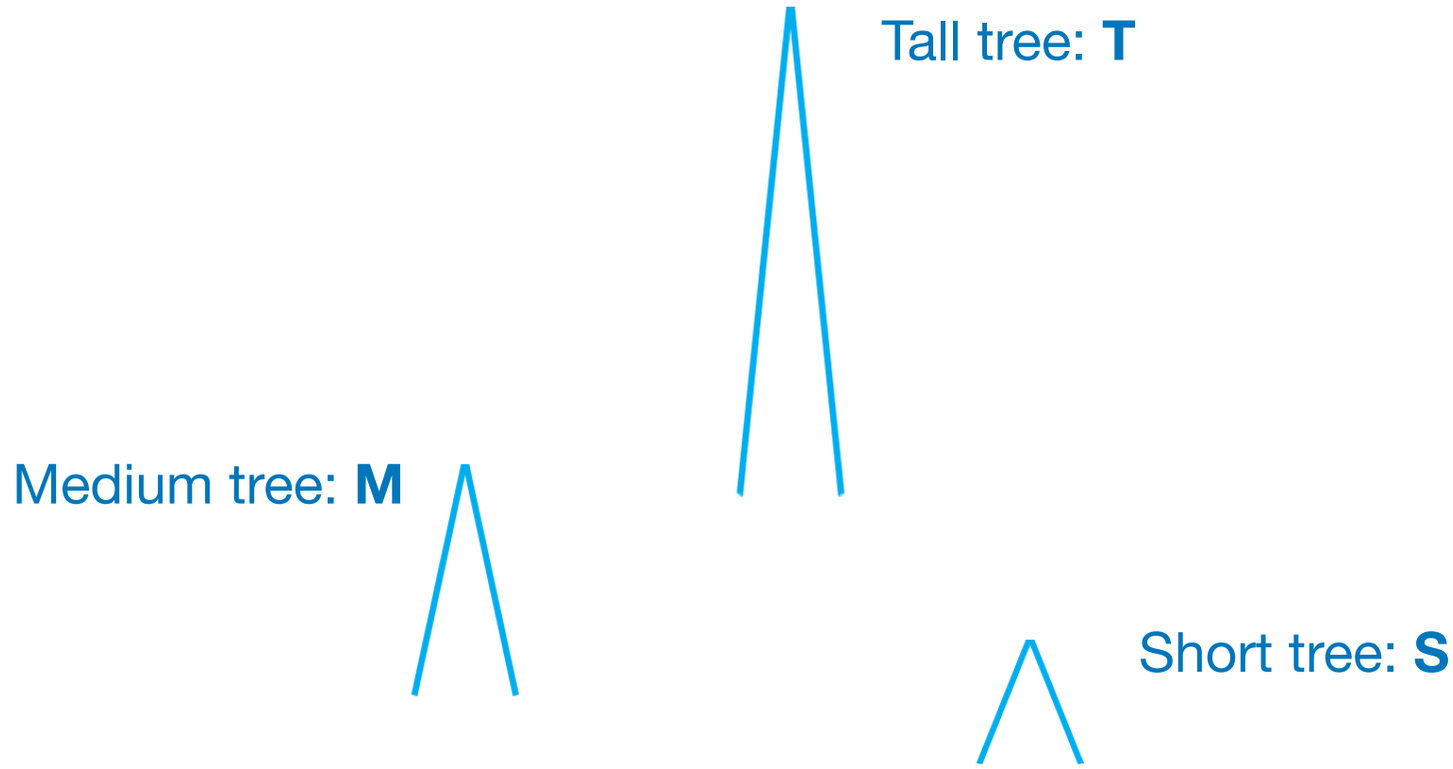
Which segment has the longest TMRCA?



- Why?

Tree heights and sequence differences

Observed data:
Homozygote: 0
Heterozygote: **1**



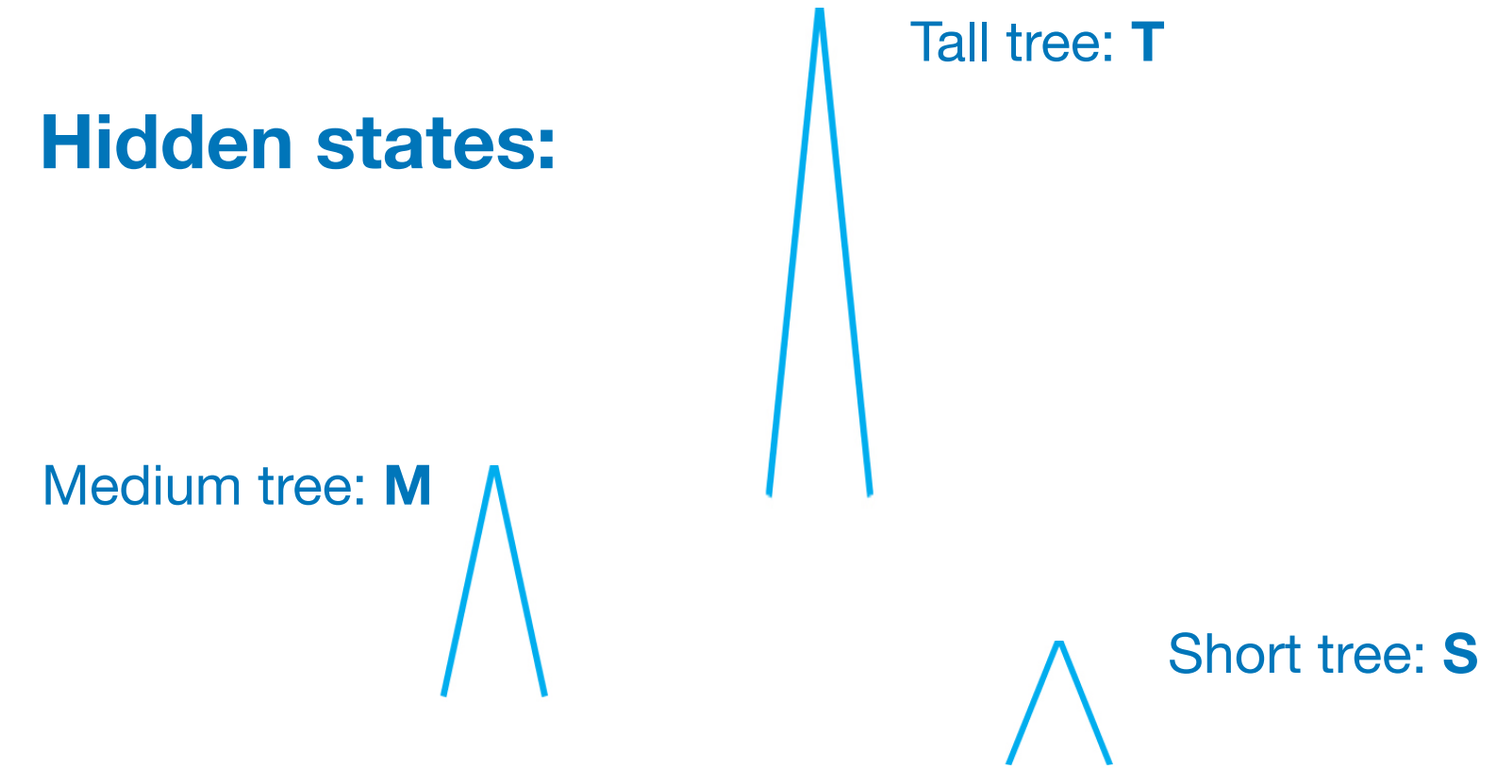
0 0 0 0 **1** 0 0 **1** **1** 0

Hidden Markov model

Tree heights and sequence differences

Observed data:
Homozygote: 0
Heterozygote: **1**

Hidden states:



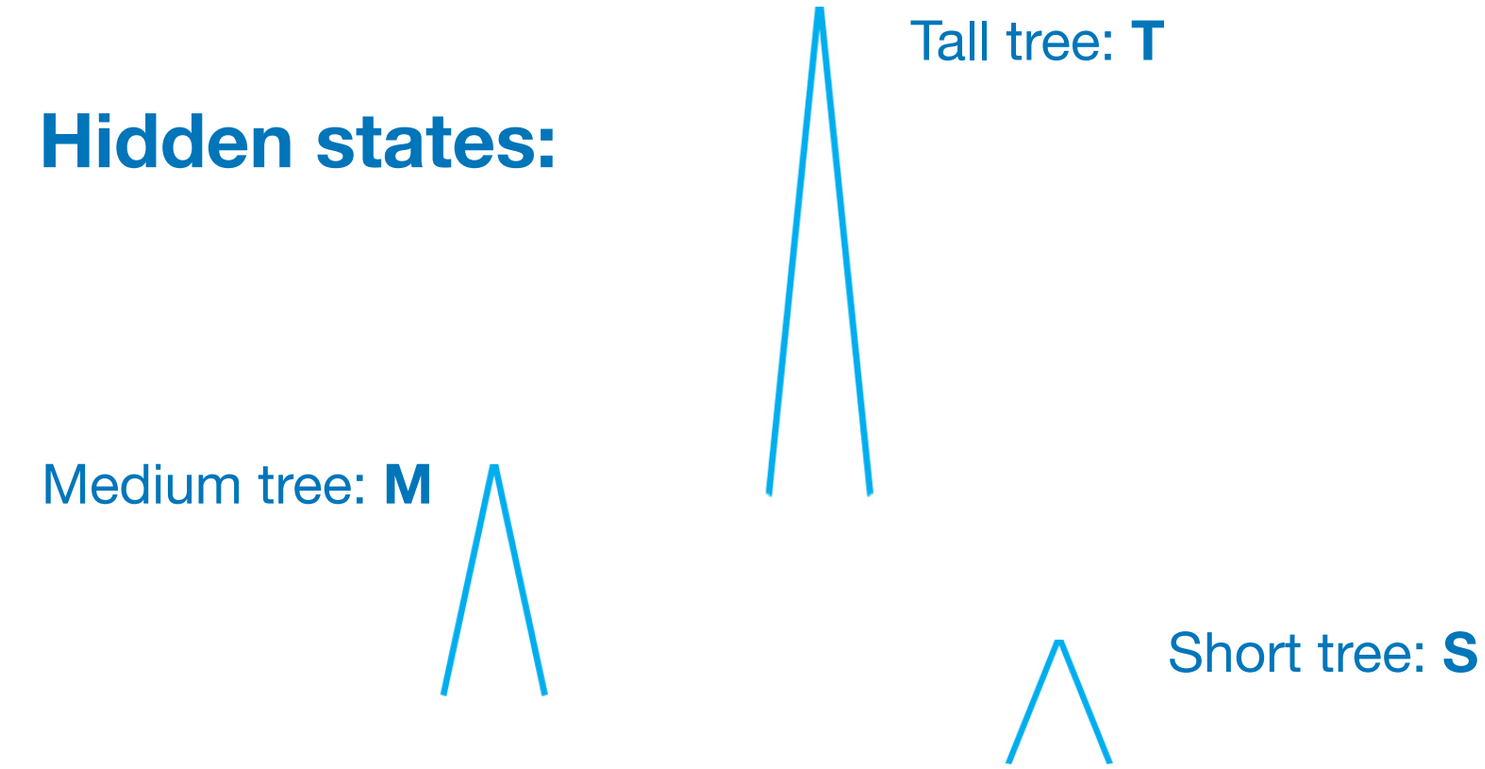
0 0 0 0 **1** 0 0 **1** **1** 0

Hidden Markov model

Tree heights and sequence differences

Observed data:
Homozygote: 0
Heterozygote: **1**

Hidden states:

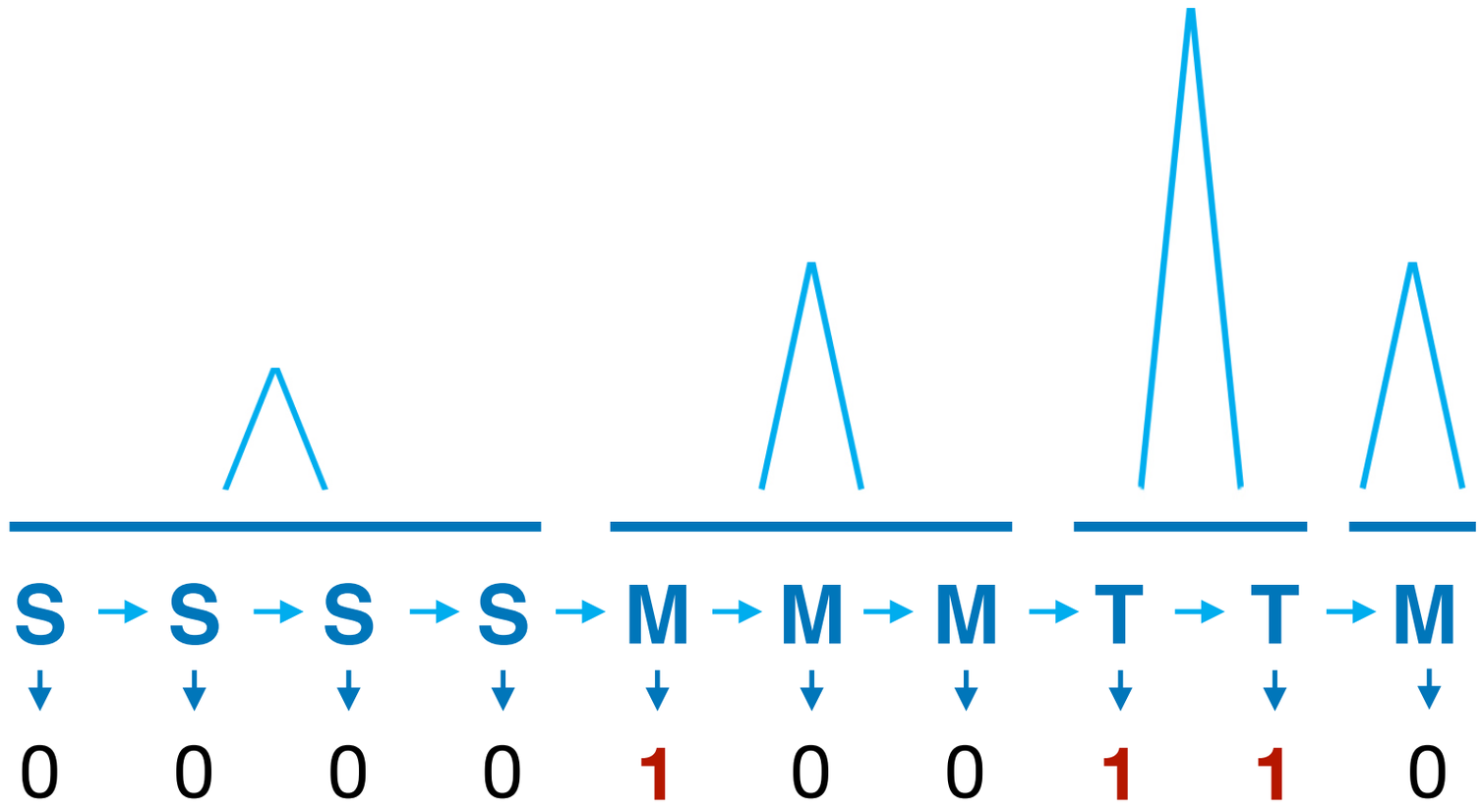


One possible hidden state path:

S	S	S	S	M	M	M	T	T	M
↓	↓	↓	↓	↓	↓	↓	↓	↓	↓
0	0	0	0	**1**	0	0	**1**	**1**	0

Hidden Markov model

Tree heights and sequence differences



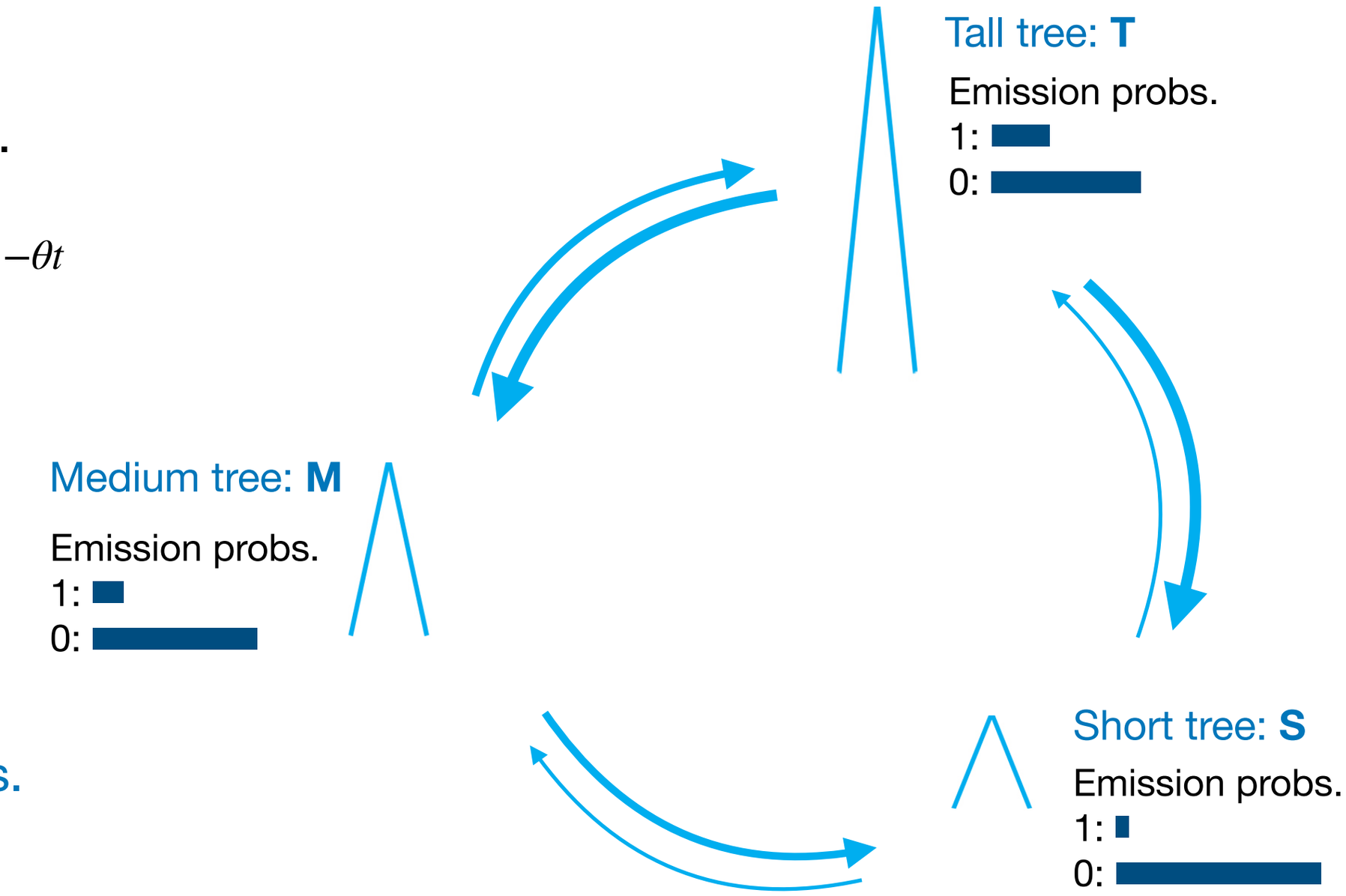
Hidden Markov model

Tree heights and sequence differences

Emission probs.

$$e(1 \mid t) = e^{-\theta t}$$

$$e(0 \mid t) = 1 - e^{-\theta t}$$



Transition probs.

$$p(t \mid s) = \ldots$$

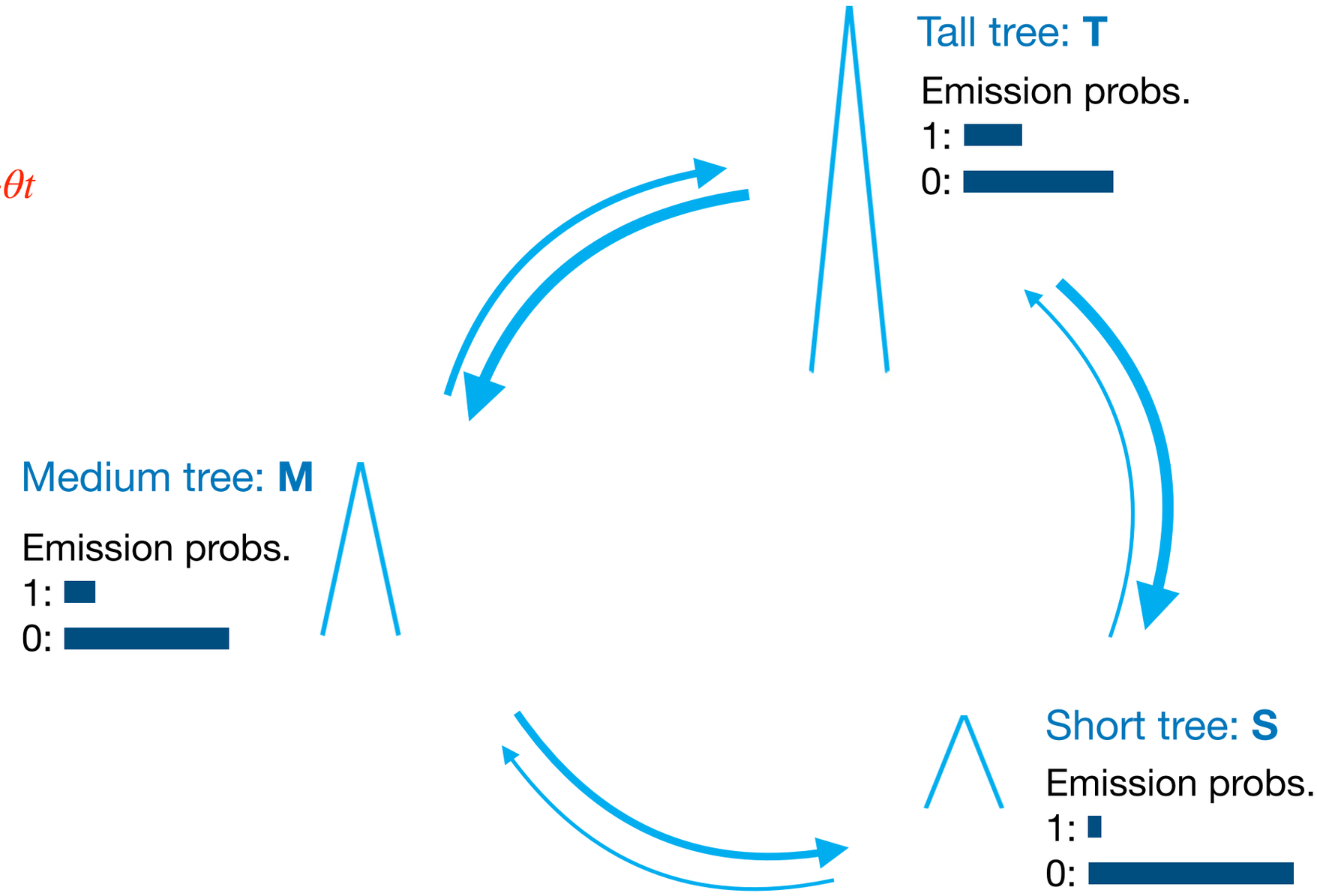
Hidden Markov model

Tree heights and sequence differences

Emission probs.

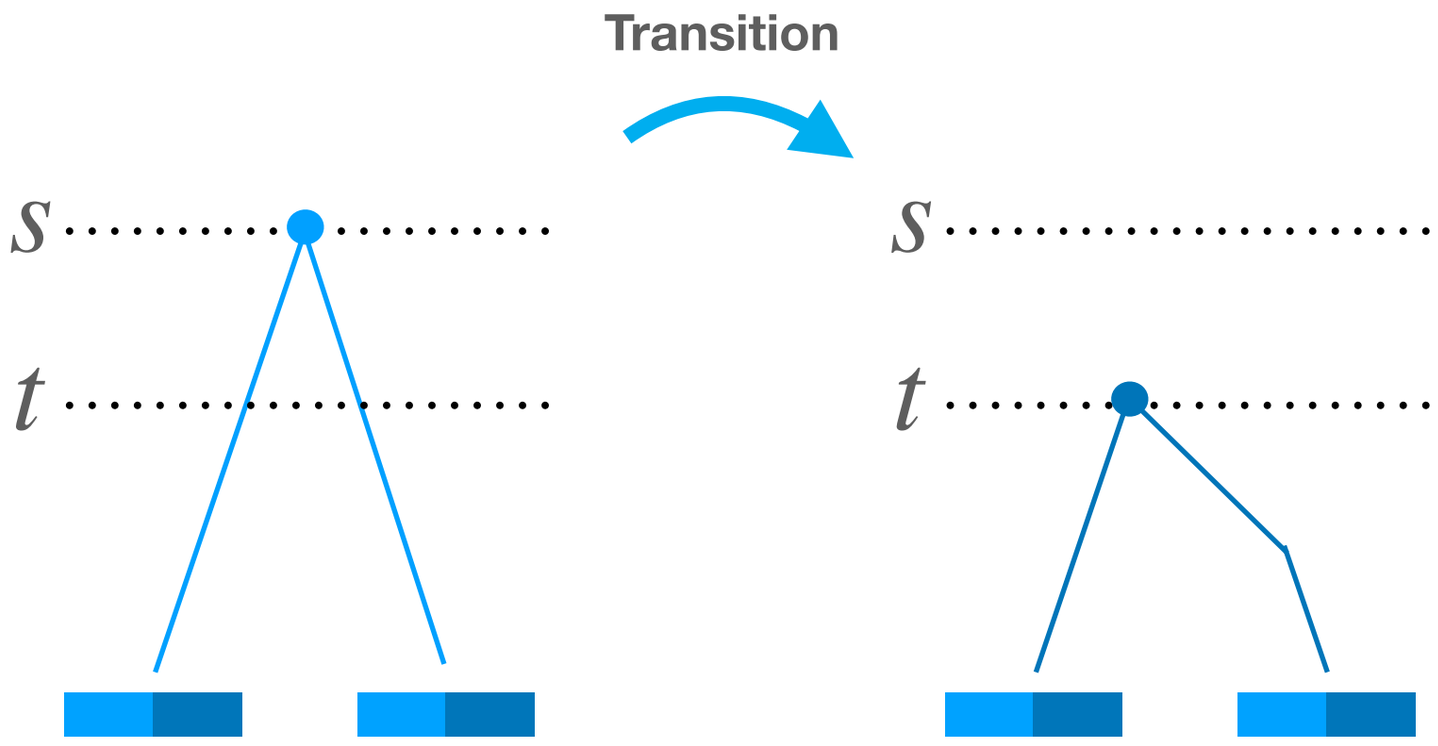
$e(1 \mid t) = 1 - e^{-\theta t}$

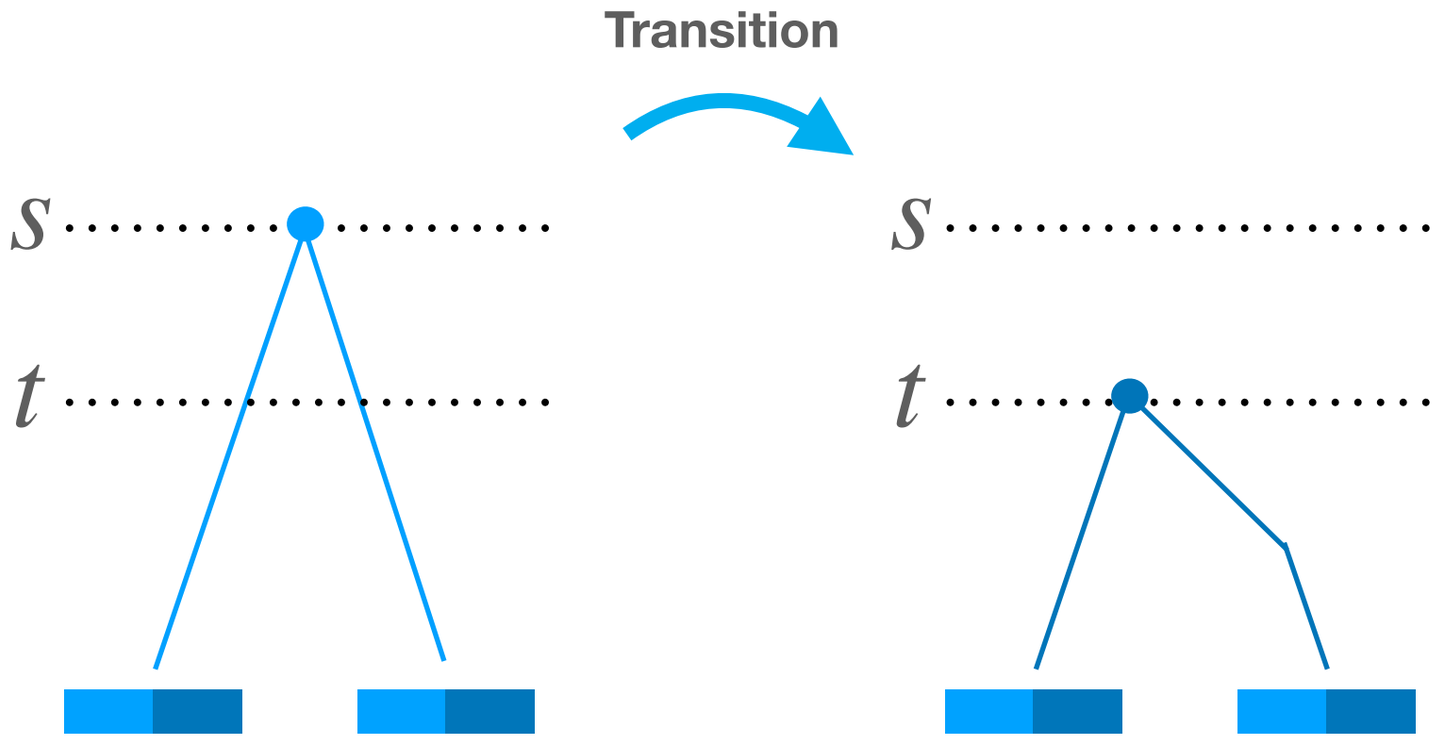
$e(0 \mid t) = e^{-\theta t}$

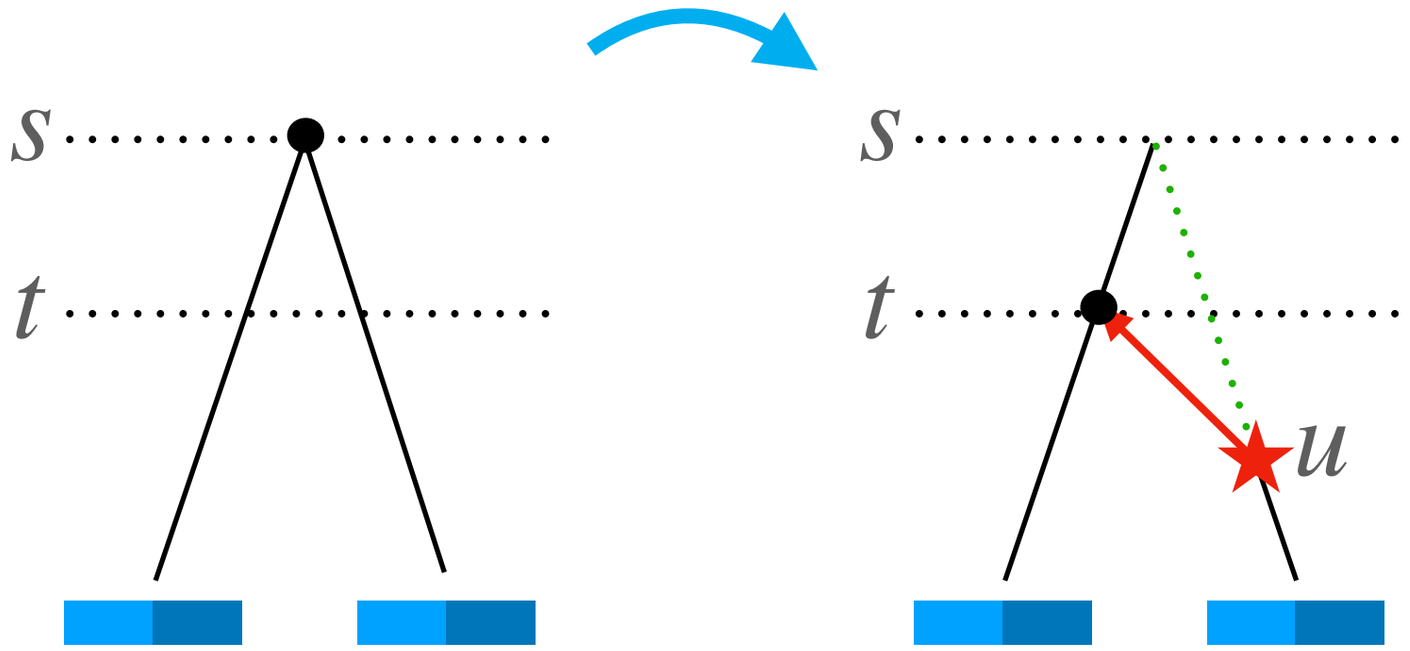


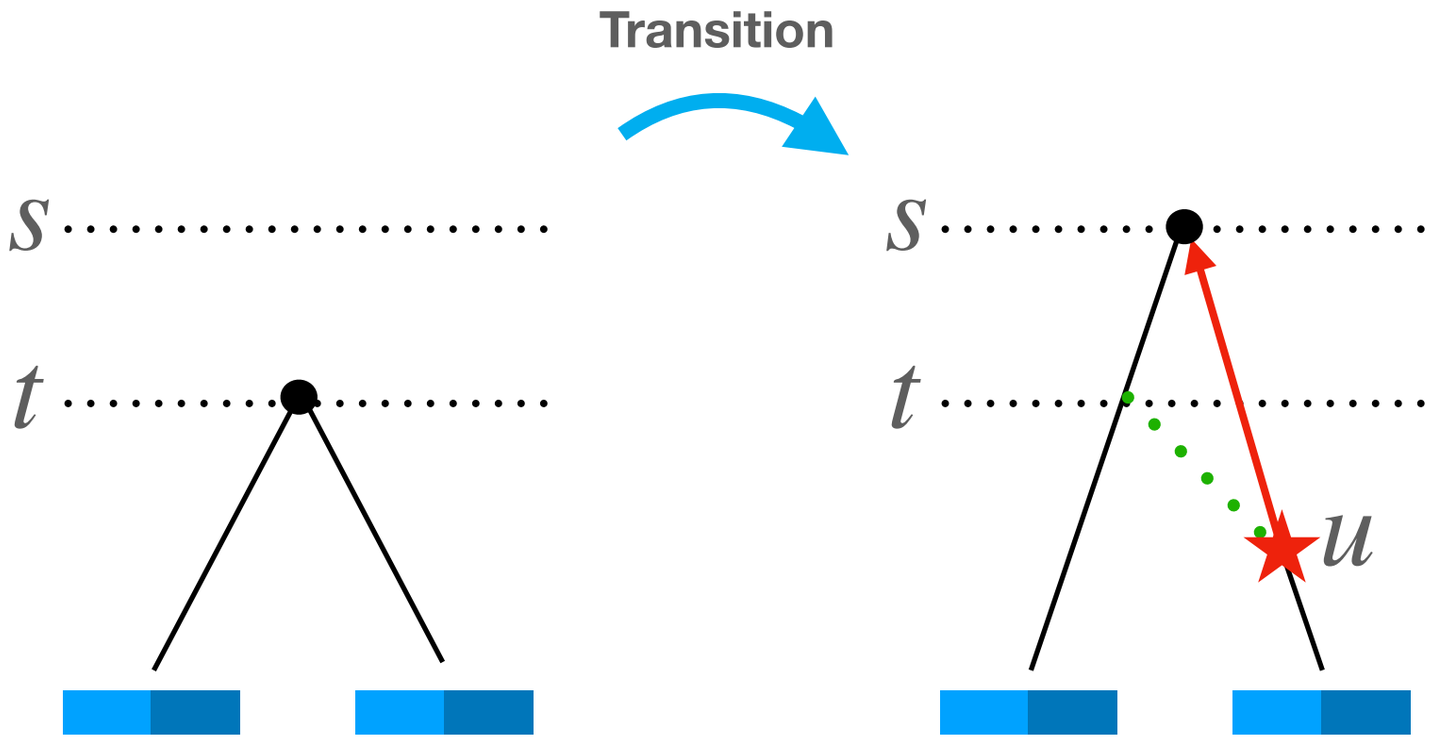
Transition probs.

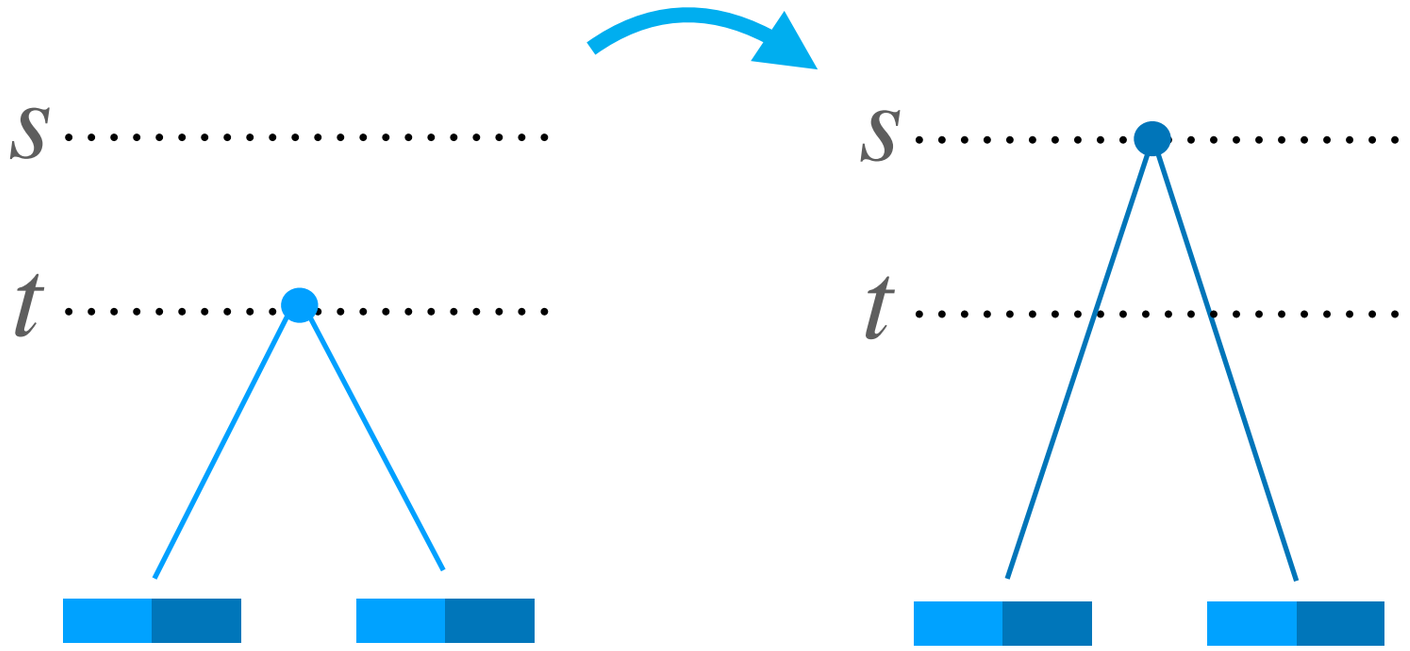
$p(t \mid s) = \ldots$

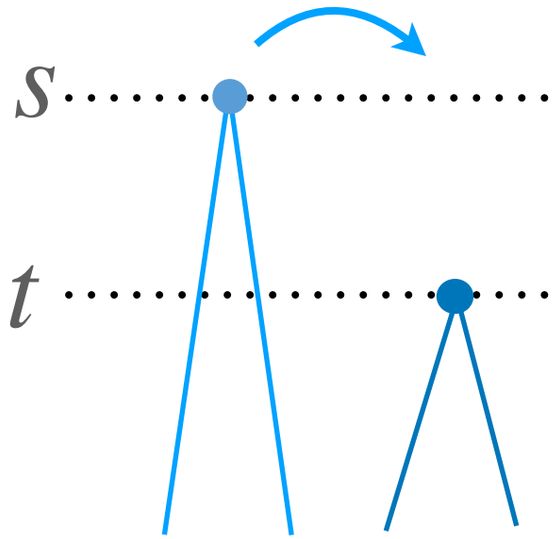

Transition
s
t
s
t

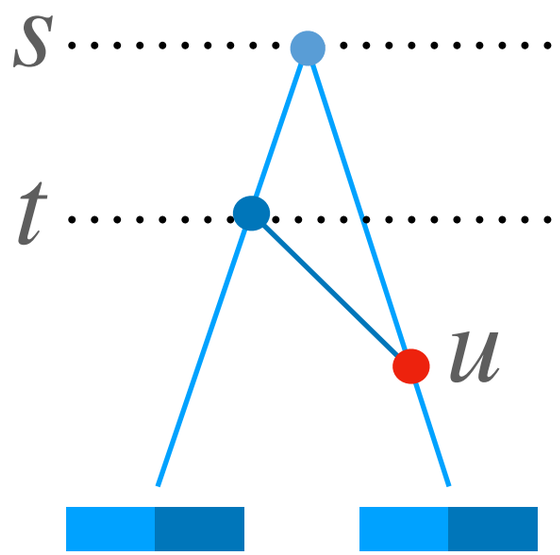

Transition
s
t
s
t


s
t
s
t
u


Transition
s
t
s
t
u


s
t
s
t





Prob. of recombination

Prob. of no recombination

1 if t = s, else 0

Transition probability from state s to state t →

$$p(t \mid s) = (1 - e^{-\rho t})\, q(t \mid s) \;+\; e^{-\rho s}\, \delta(t - s)$$

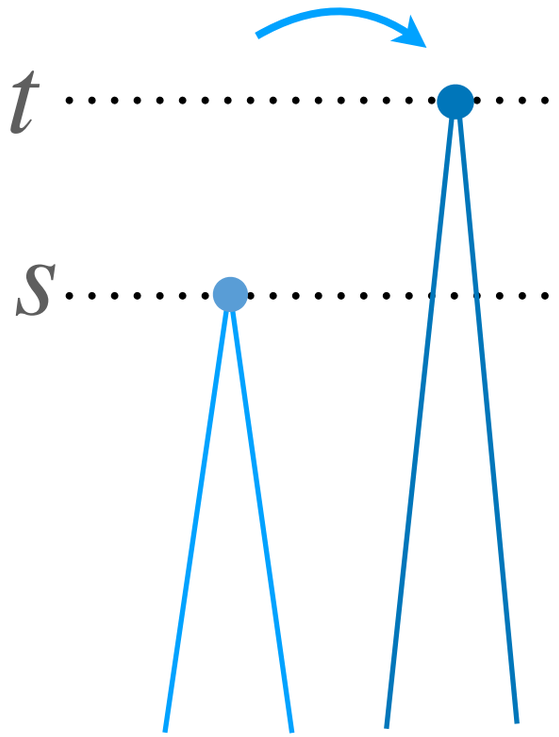
Mean over recombination times u (between 0 and t))

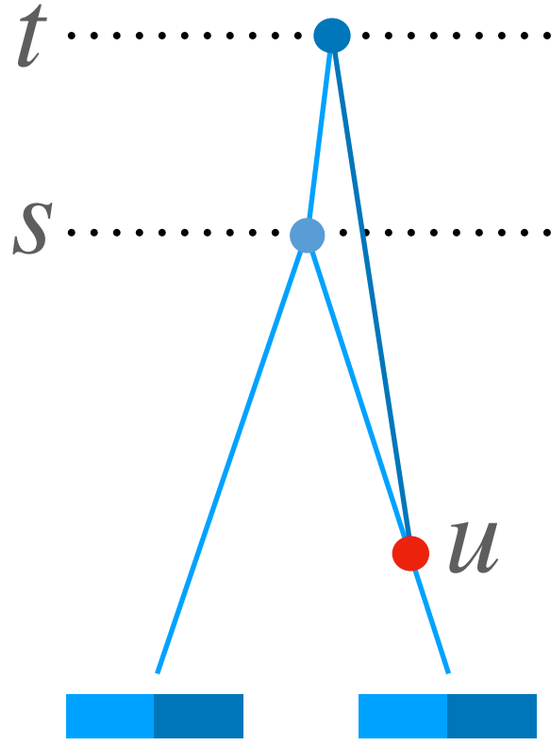
Prob of change to state t upon recombination: → $q(t \mid s)$

Prob. of coalescing at time t

Prob. of not coalescing between time u and t

Ne (scaled) at time t





Prob. of recombination

Prob. of no recombination

1 if t = s

else 0

Transition probability from state s to state t →

$$p(t \mid s) = (1 - e^{-\rho t})\, q(t \mid s) \;+\; e^{-\rho s}\, \delta(t - s)$$

Mean over recombination times u (between 0 and t))

Prob of change to state t upon recombination: → $q(t \mid s)$

Prob. of coalescing at time t

Prob. of not coalescing between time u and t

Ne (scaled) at time t

$$p(t \mid s) = (1 - e^{-\rho t})\ q(t \mid s)\ \ +\ \ e^{-\rho s}\,\delta(t - s)$$

$$q(t \mid s) = \frac{1}{\lambda(t)} \int_0^{min(s,t)} \frac{1}{s} \times e^{-\int_u^t \frac{1}{\lambda(v)} dv} du$$

Prob. of coalescing at time t

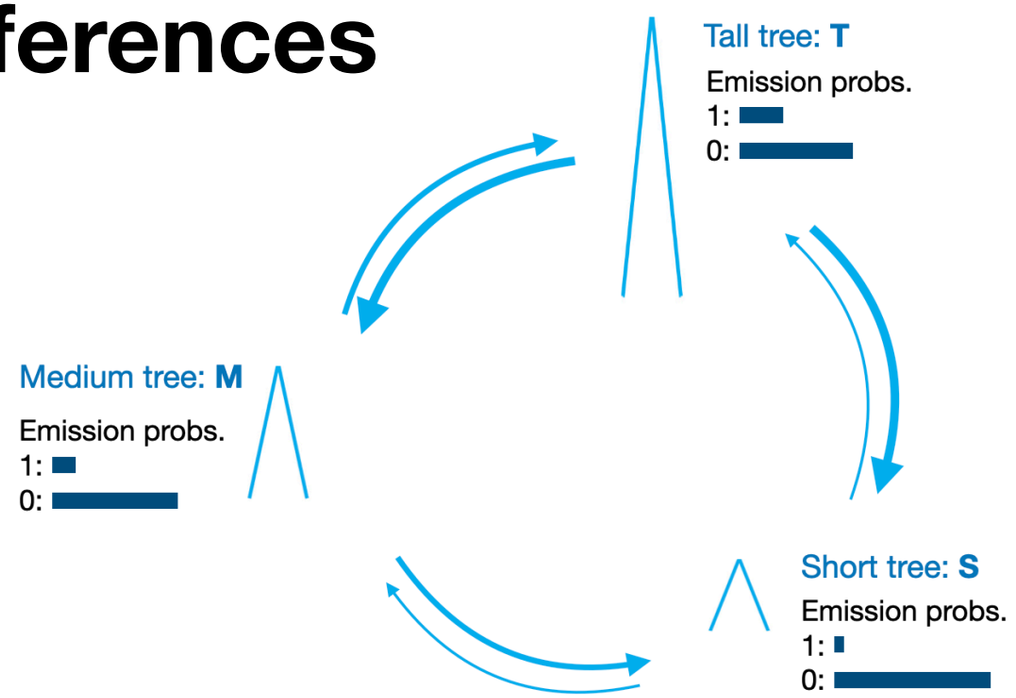
Prob. of not coalescing between time u and t

$$\lambda(t) = \frac{Ne(t)}{N_0}$$

Ne (scaled) at time t

Hidden Markov model

Tree heights and sequence differences



One possible hidden state path:

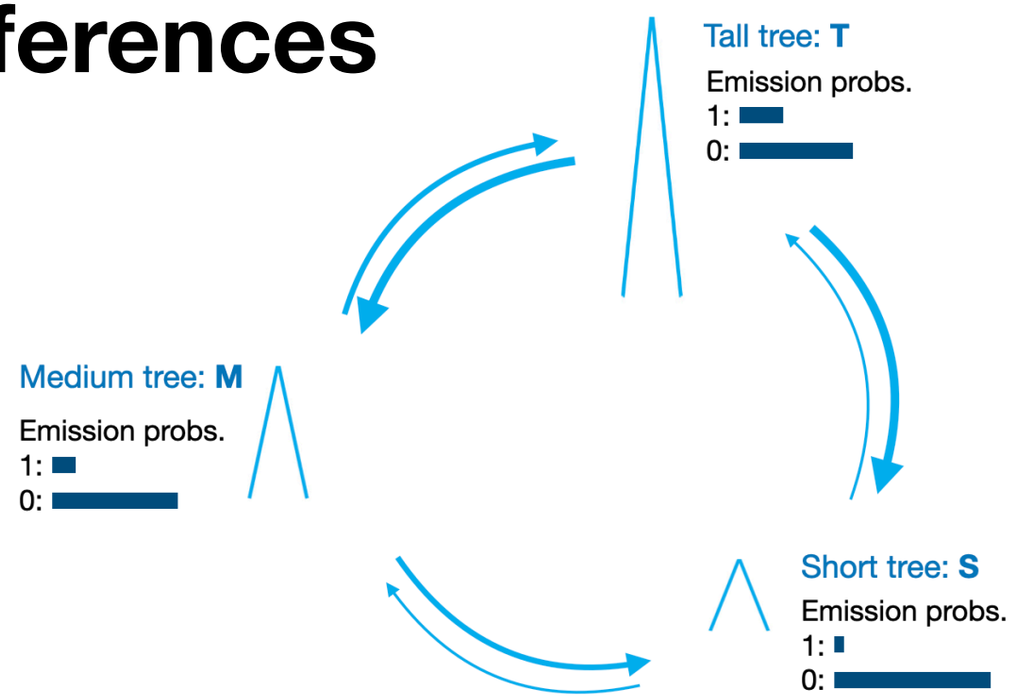
S	S	S	S	M	M	M	T	T	M
↓	↓	↓	↓	↓	↓	↓	↓	↓	↓
0	0	0	0	1	0	0	1	1	0

Another possible hidden state path:

M	M	S	S	S	S	S	T	T	T
↓	↓	↓	↓	↓	↓	↓	↓	↓	↓
0	0	0	0	1	0	0	1	1	0

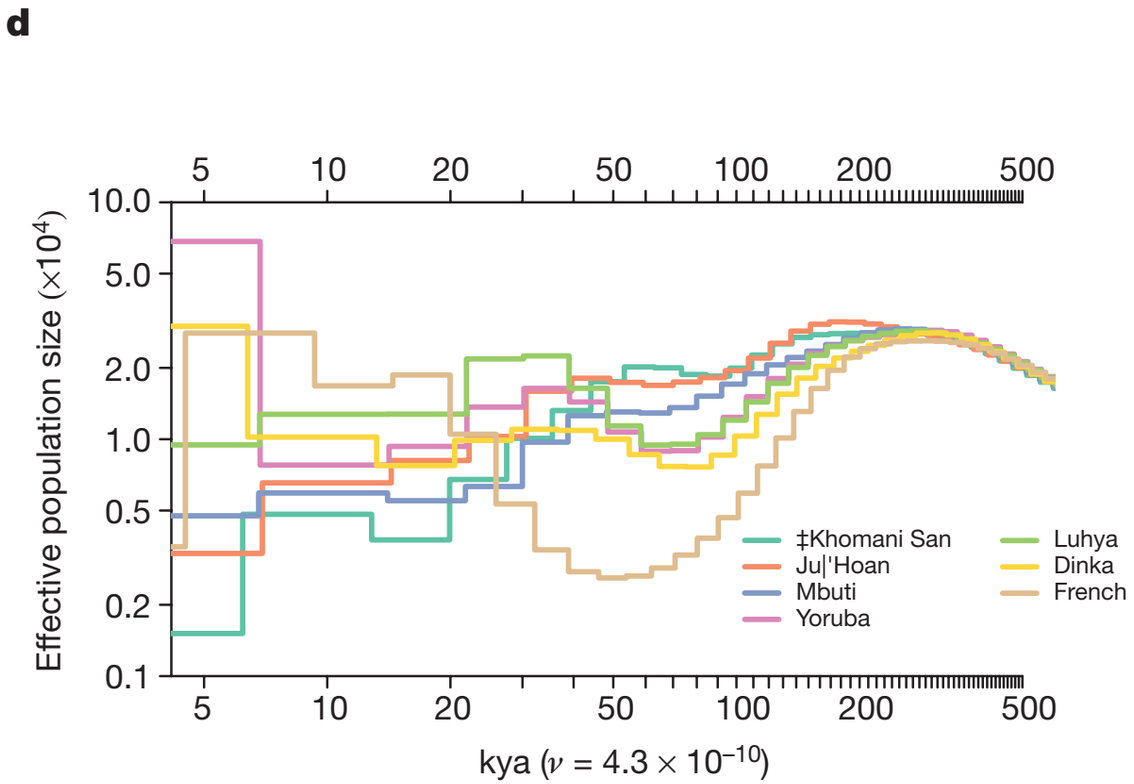
Hidden Markov model

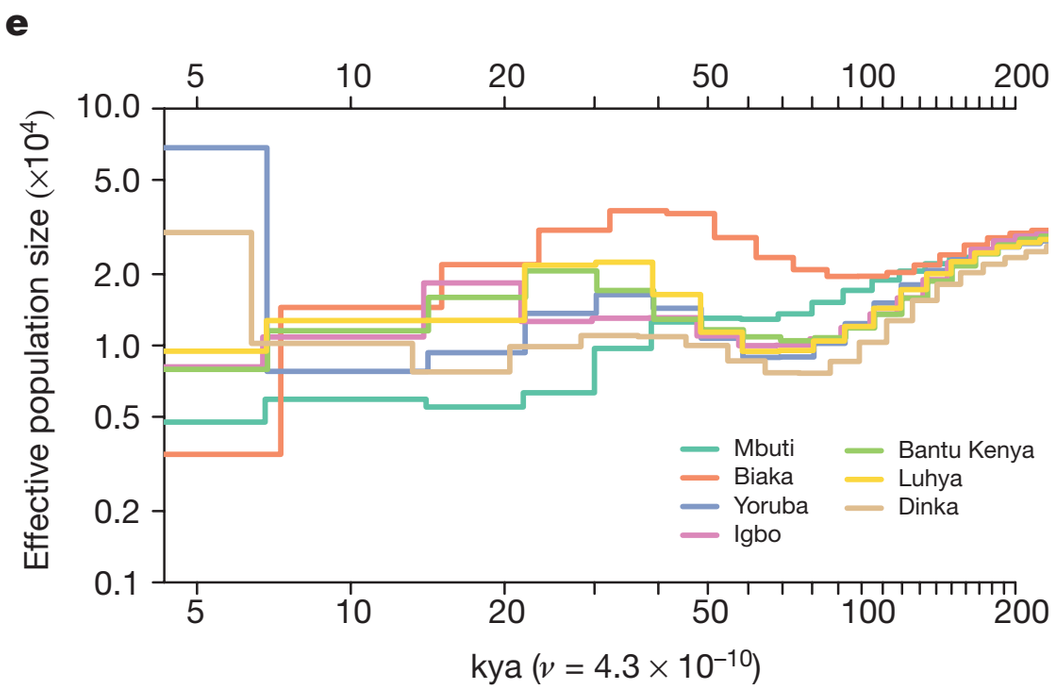
Tree heights and sequence differences

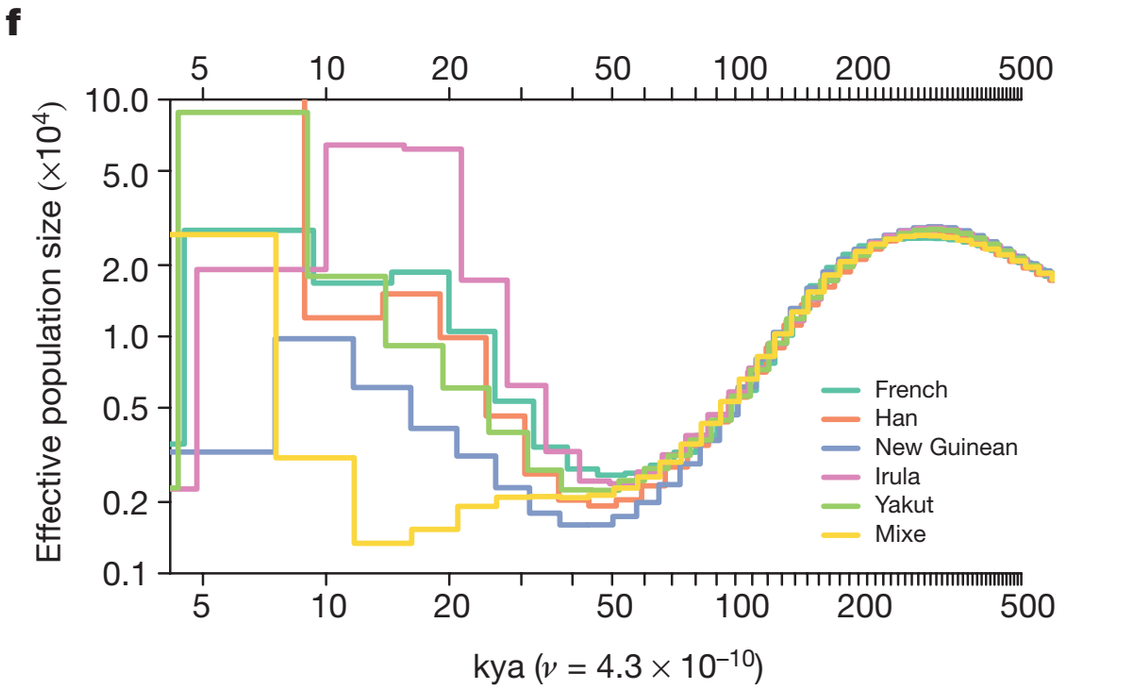


- How do we find the past population sizes that makes our data most likely, across all possible hidden state paths?

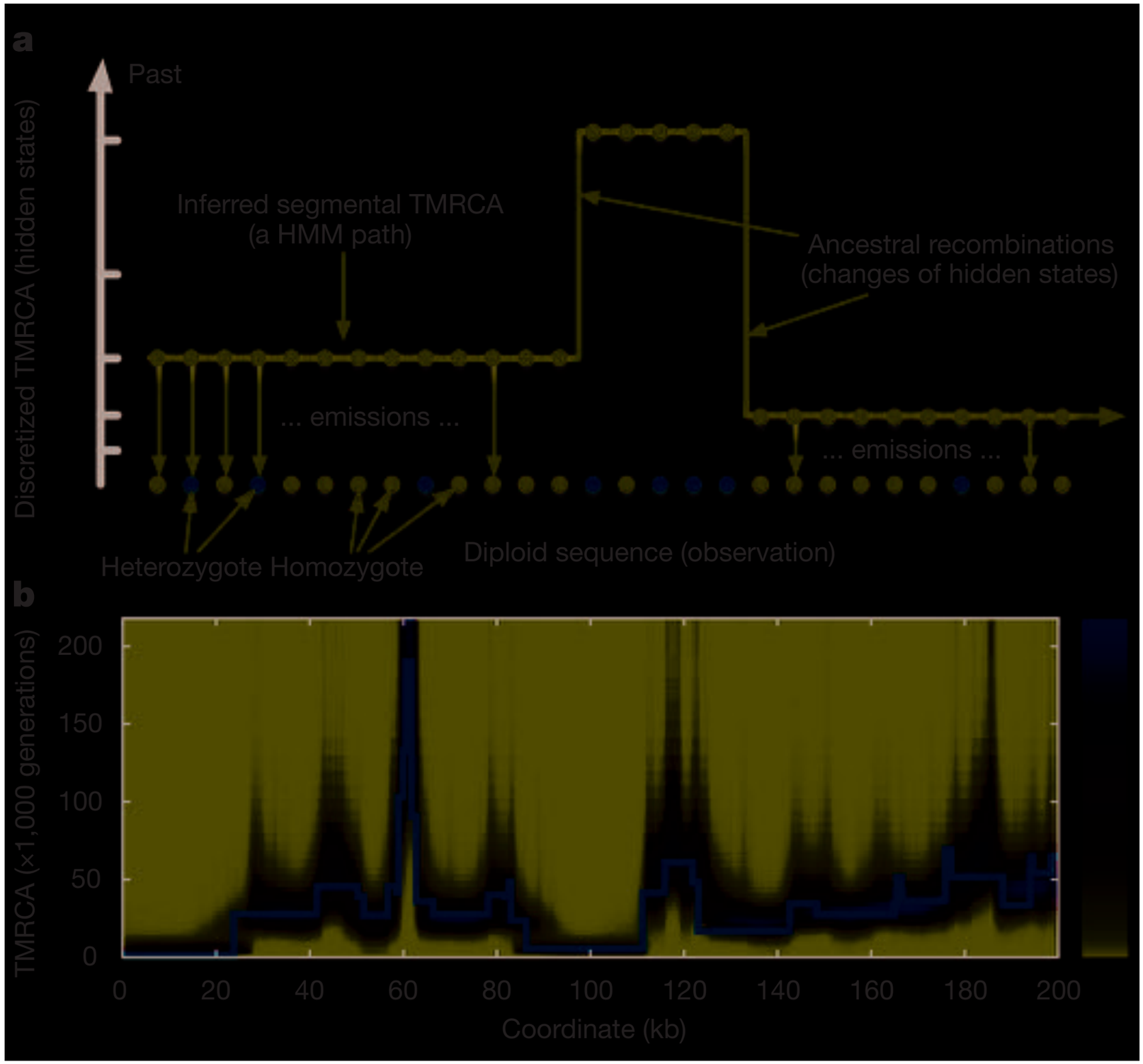
SGDP paper


d
5
10
20
50
100
200
500
10.0
5.0
2.0
1.0
0.5
0.2
0.1
Effective population size (×10⁴)
‡Khomani San
Ju|'Hoan
Mbuti
Yoruba
Luhya
Dinka
French
kya (ν = 4.3 × 10⁻¹⁰)


e
5
10
20
50
100
200
10.0
5.0
2.0
1.0
0.5
0.2
0.1
Effective population size (×10⁴)
Mbuti
Biaka
Yoruba
Igbo
Bantu Kenya
Luhya
Dinka
kya (ν = 4.3 × 10⁻¹⁰)


f
5
10
20
50
100
200
500
10.0
5.0
2.0
1.0
0.5
0.2
0.1
Effective population size (×10⁴)
French
Han
New Guinean
Irula
Yakut
Mixe
kya (ν = 4.3 × 10⁻¹⁰)

PSMC


a
Past
Discretized TMRCA (hidden states)
Inferred segmental TMRCA
(a HMM path)
Ancestral recombinations
(changes of hidden states)
... emissions ...
... emissions ...
Heterozygote Homozygote
Diploid sequence (observation)
b
TMRCA (×1,000 generations)
200
150
100
50
0
0
20
40
60
80
100
120
140
160
180
200
Coordinate (kb)