

A photograph of a man with a beard and short hair, wearing a dark jacket over a green and brown striped sweater. He is standing next to a human skeleton. The skeleton's arms are raised, holding onto thin vertical lines. The background is solid black.

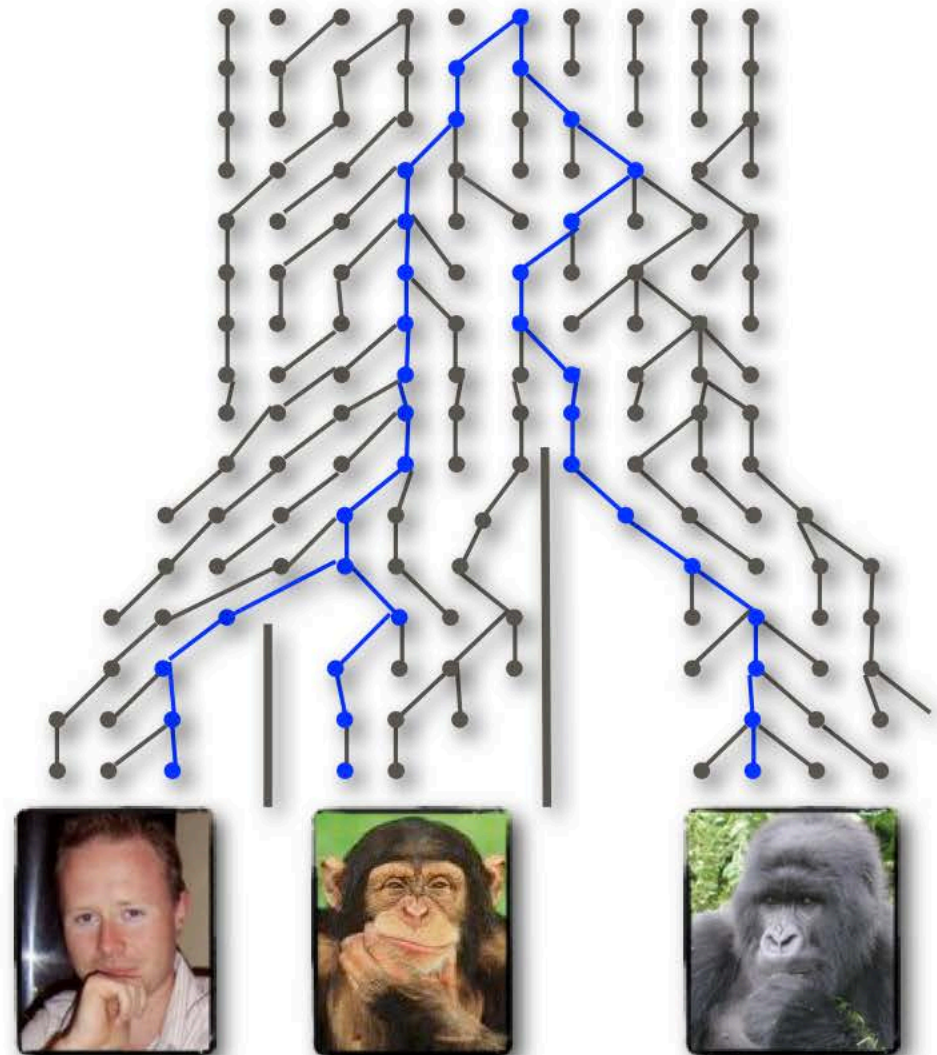
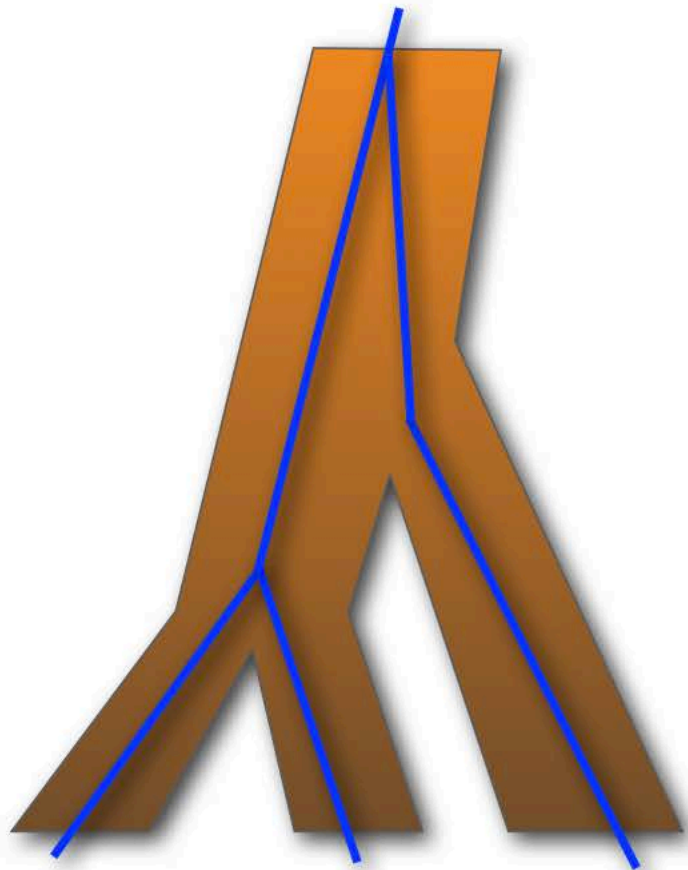
Ancestral population genomics

The sequential Markov coalescence (SMC)
and
Coalescence hidden Markov models (CoalHMMs)

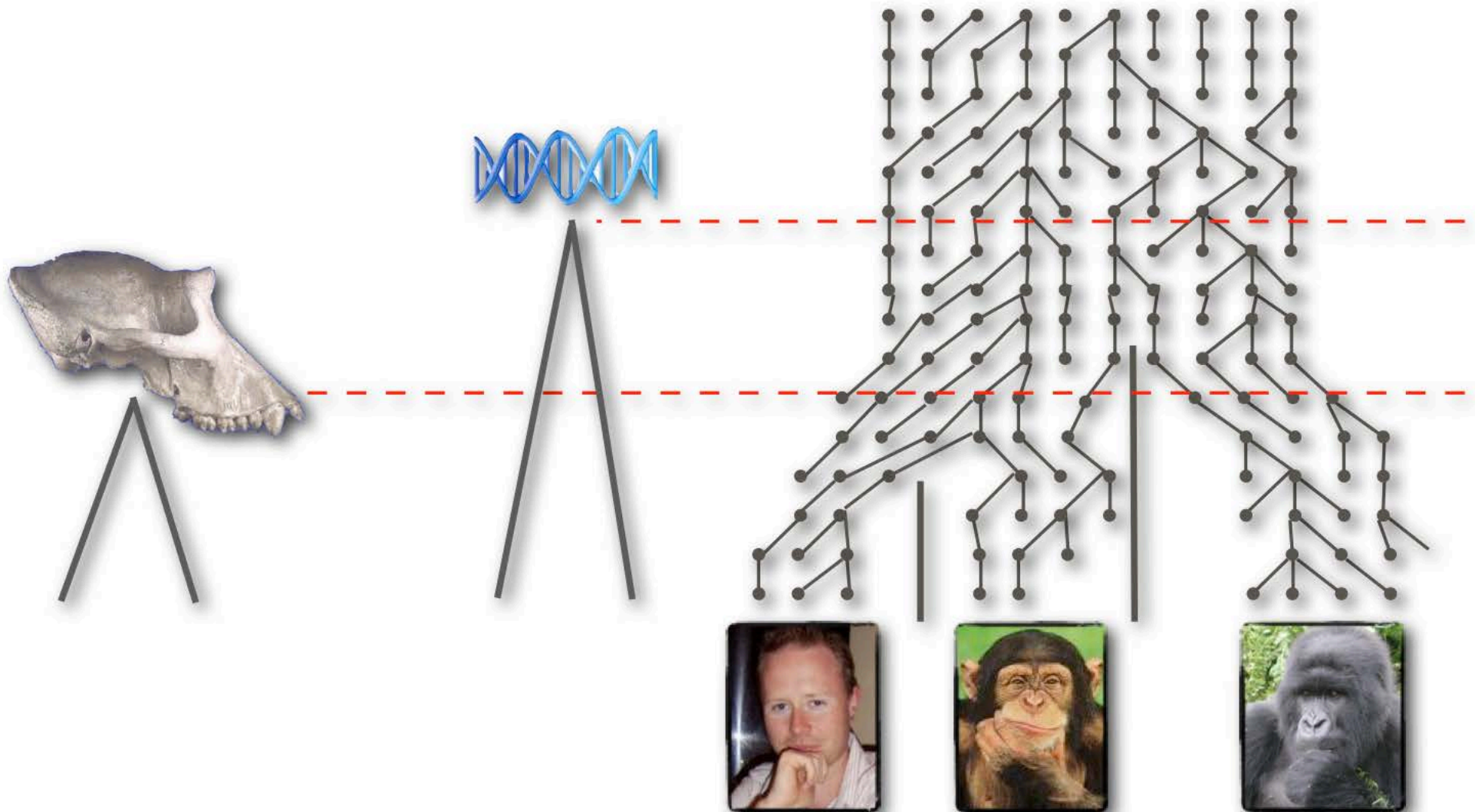
Thomas Mailund

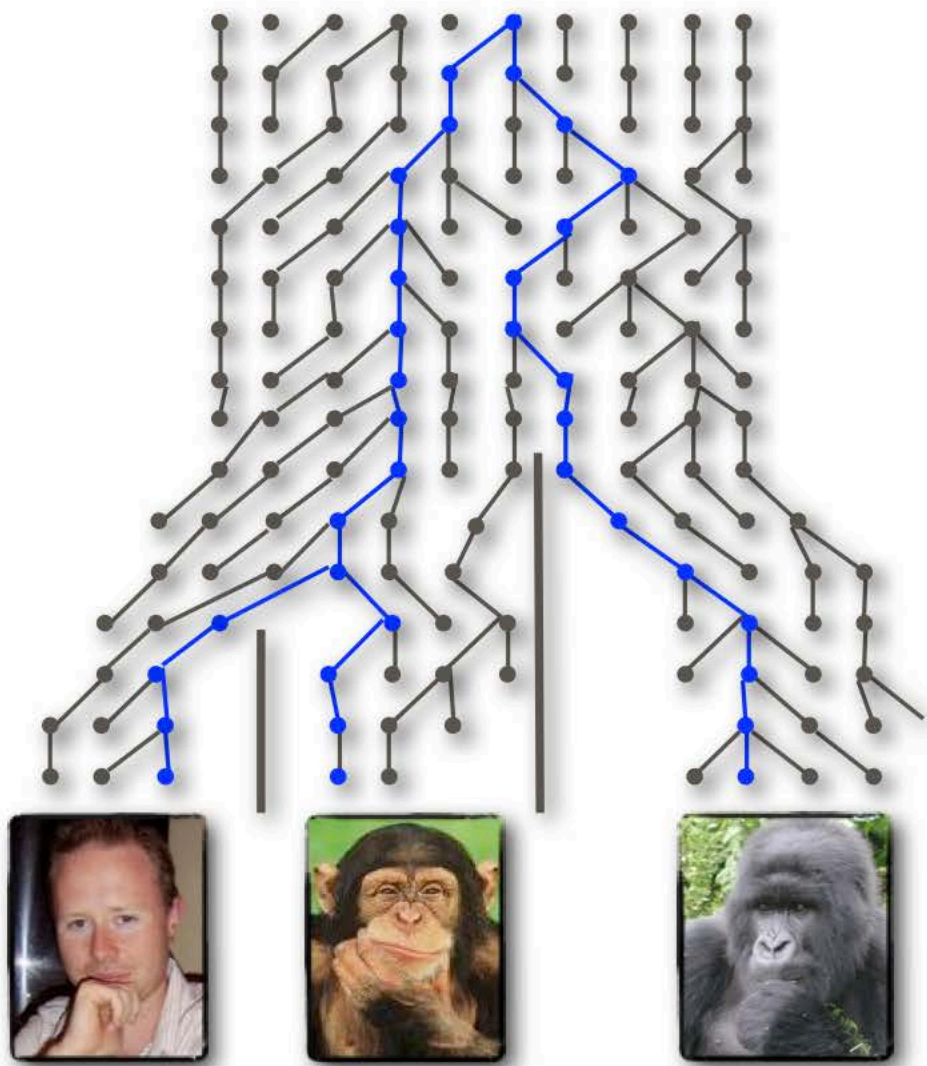
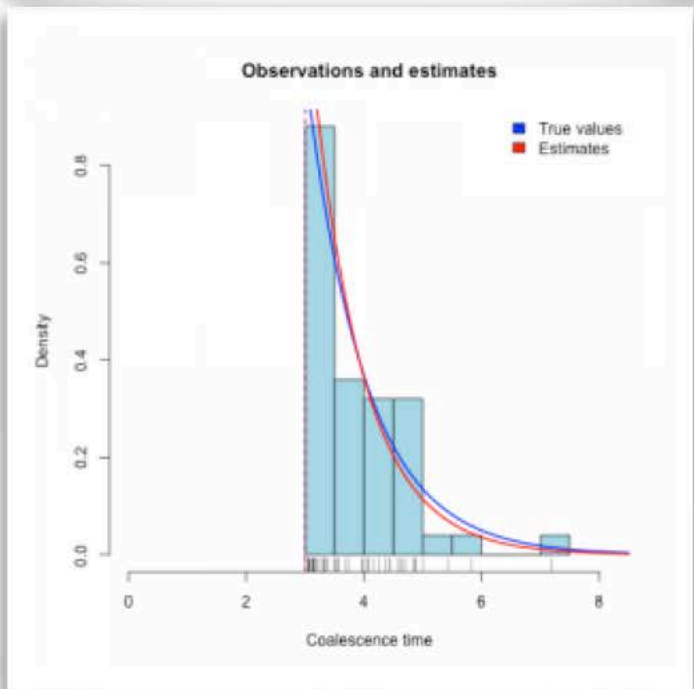
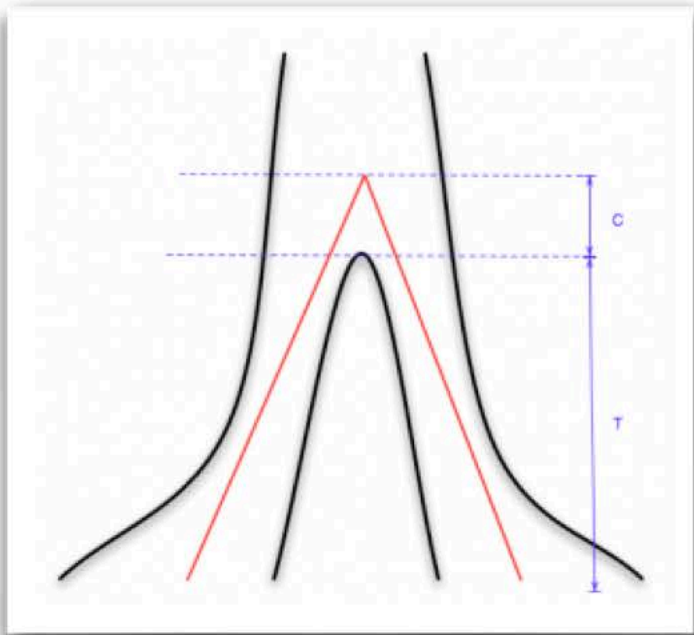
Bioinformatics Research Center
Aarhus University

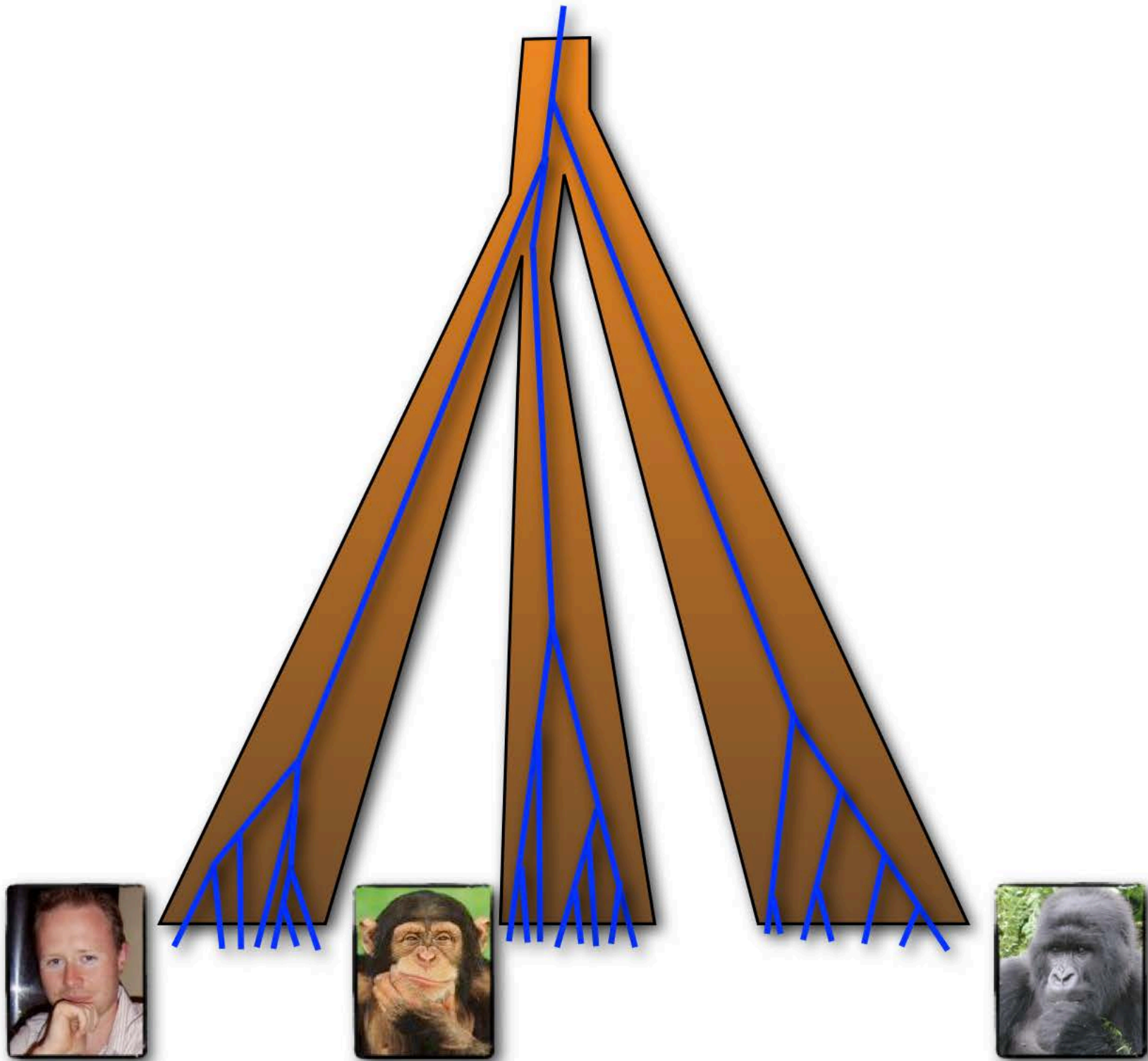
Species divergence and sequence divergence



Species divergence and sequence divergence



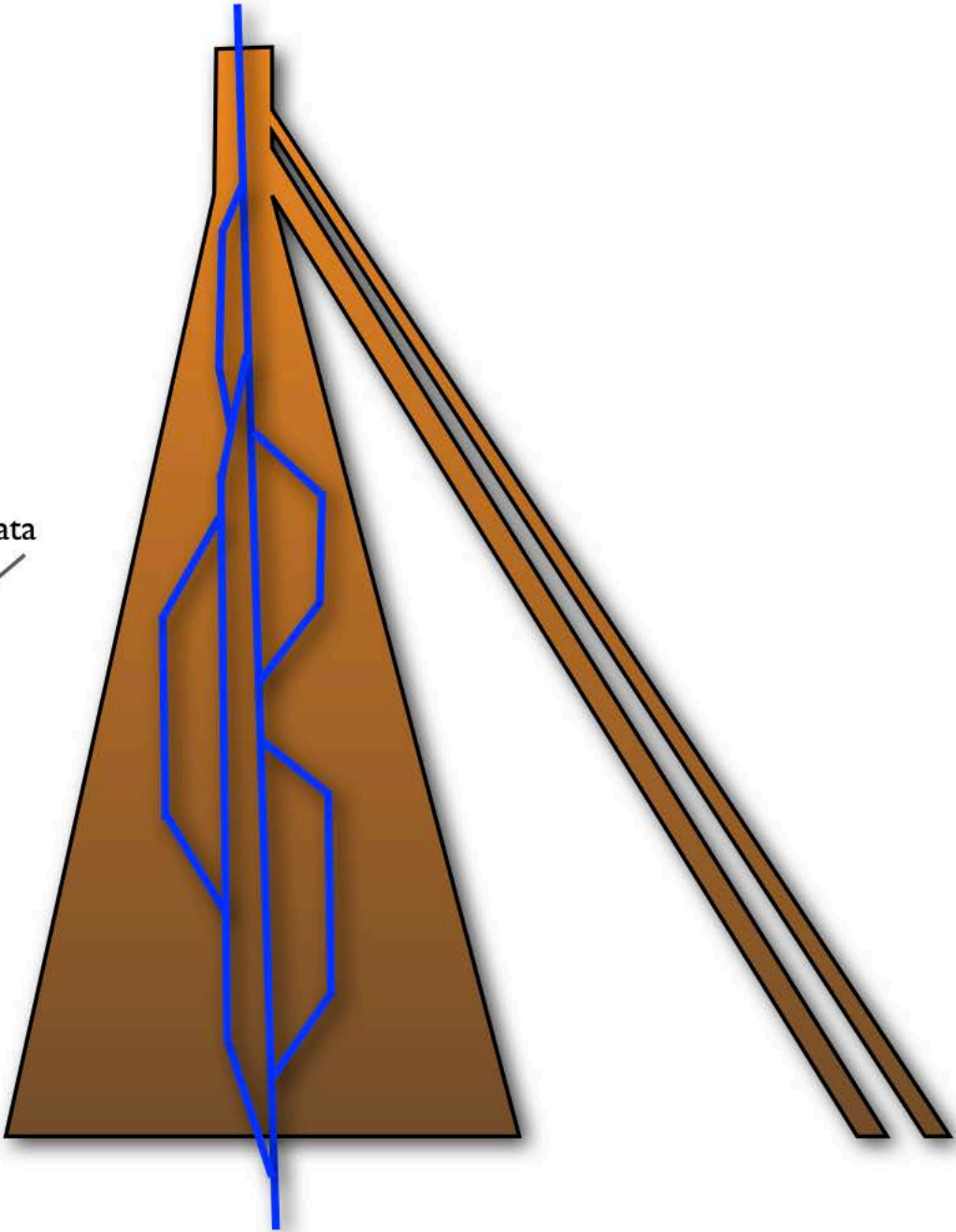


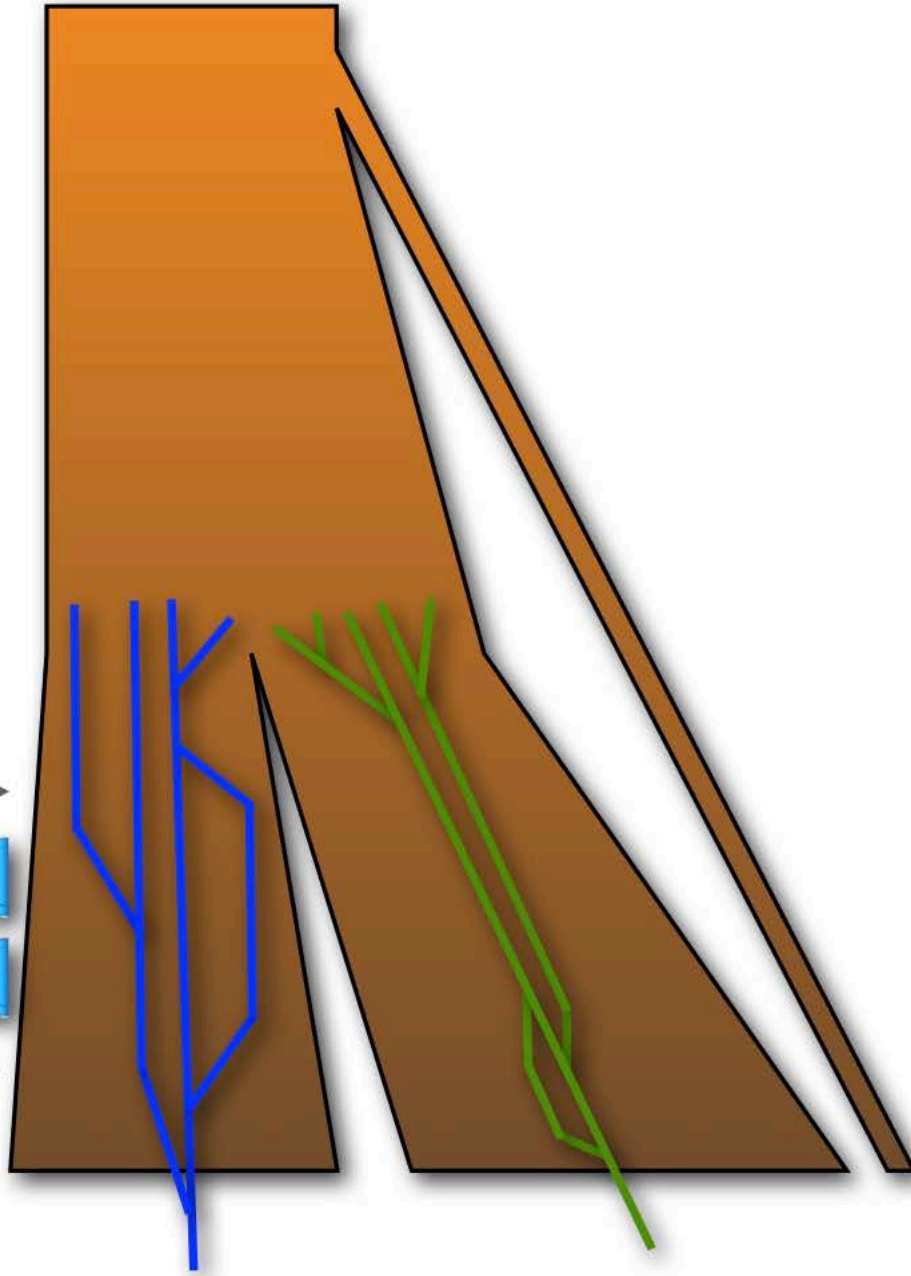
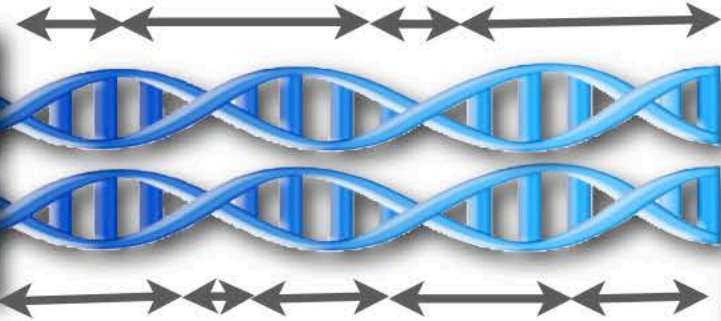


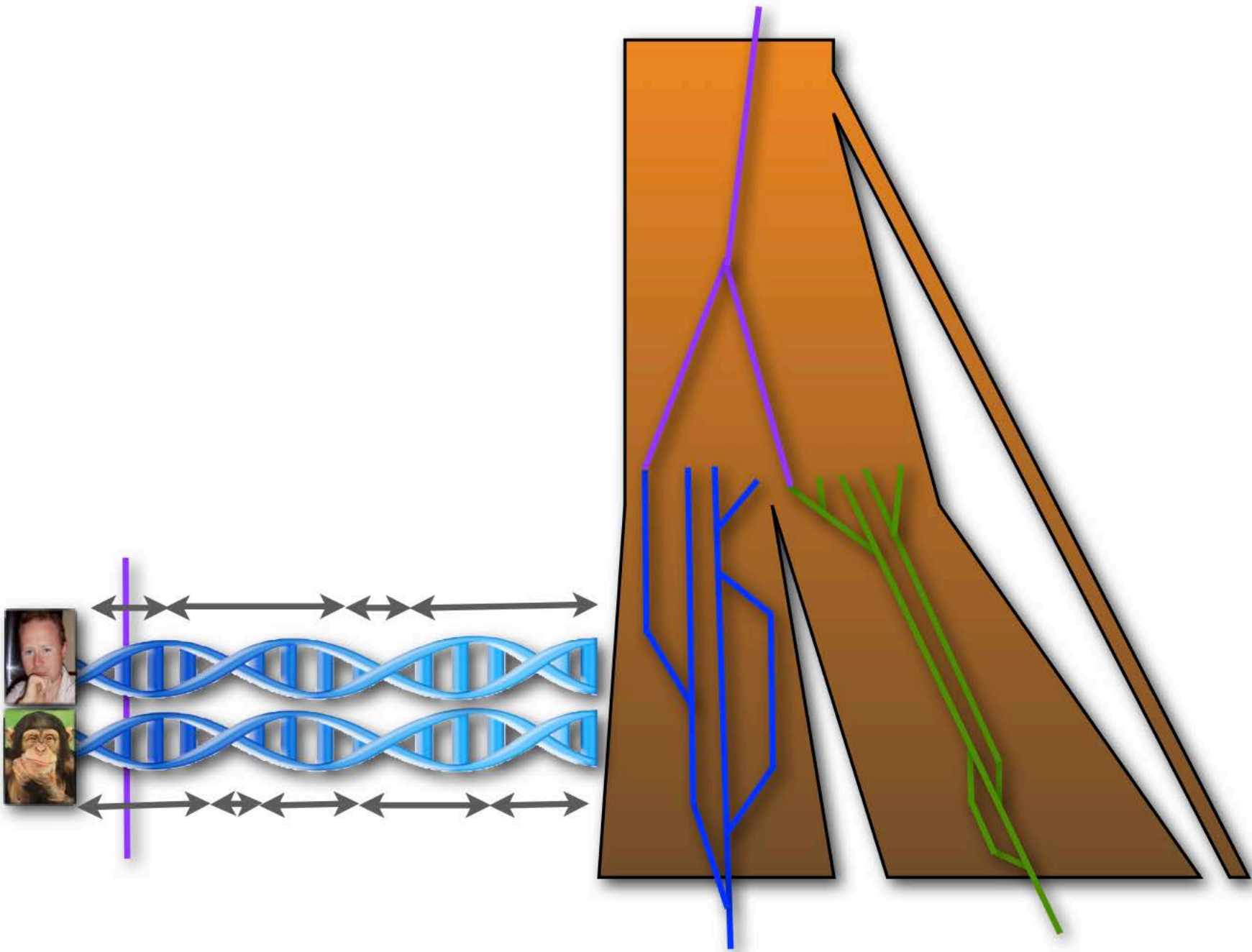
acctgatttagcata

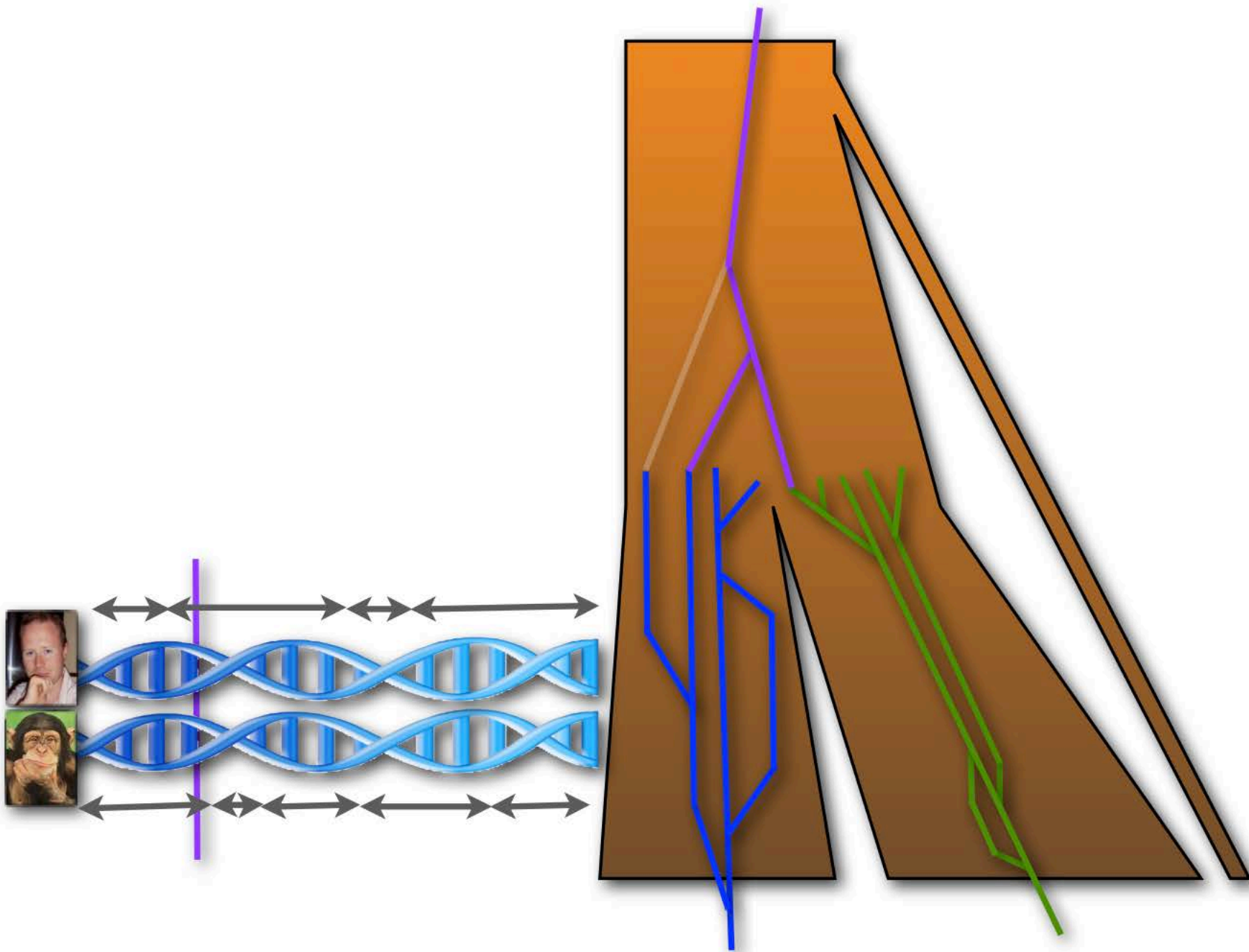
acctgatttagcata

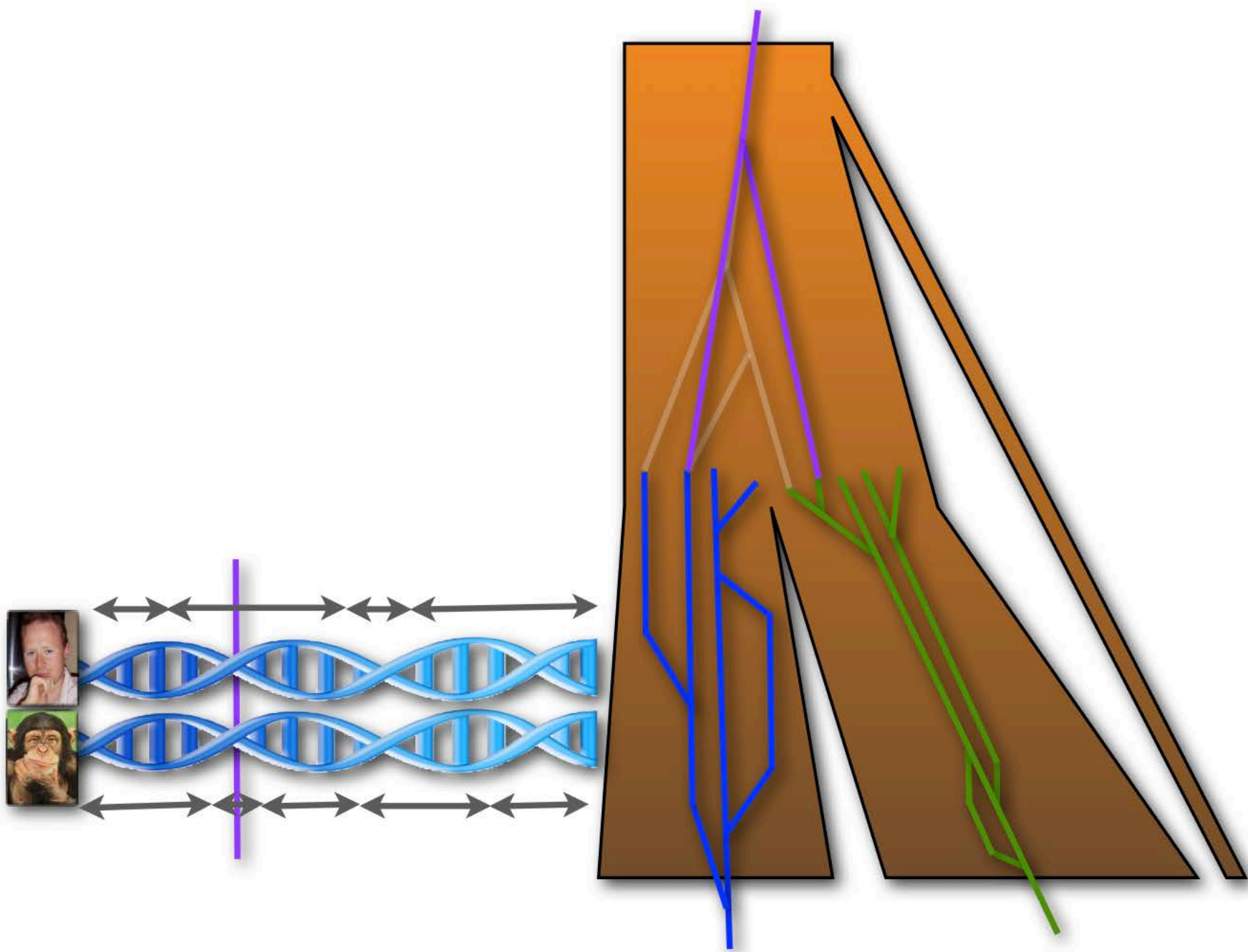
acctgatttagcata

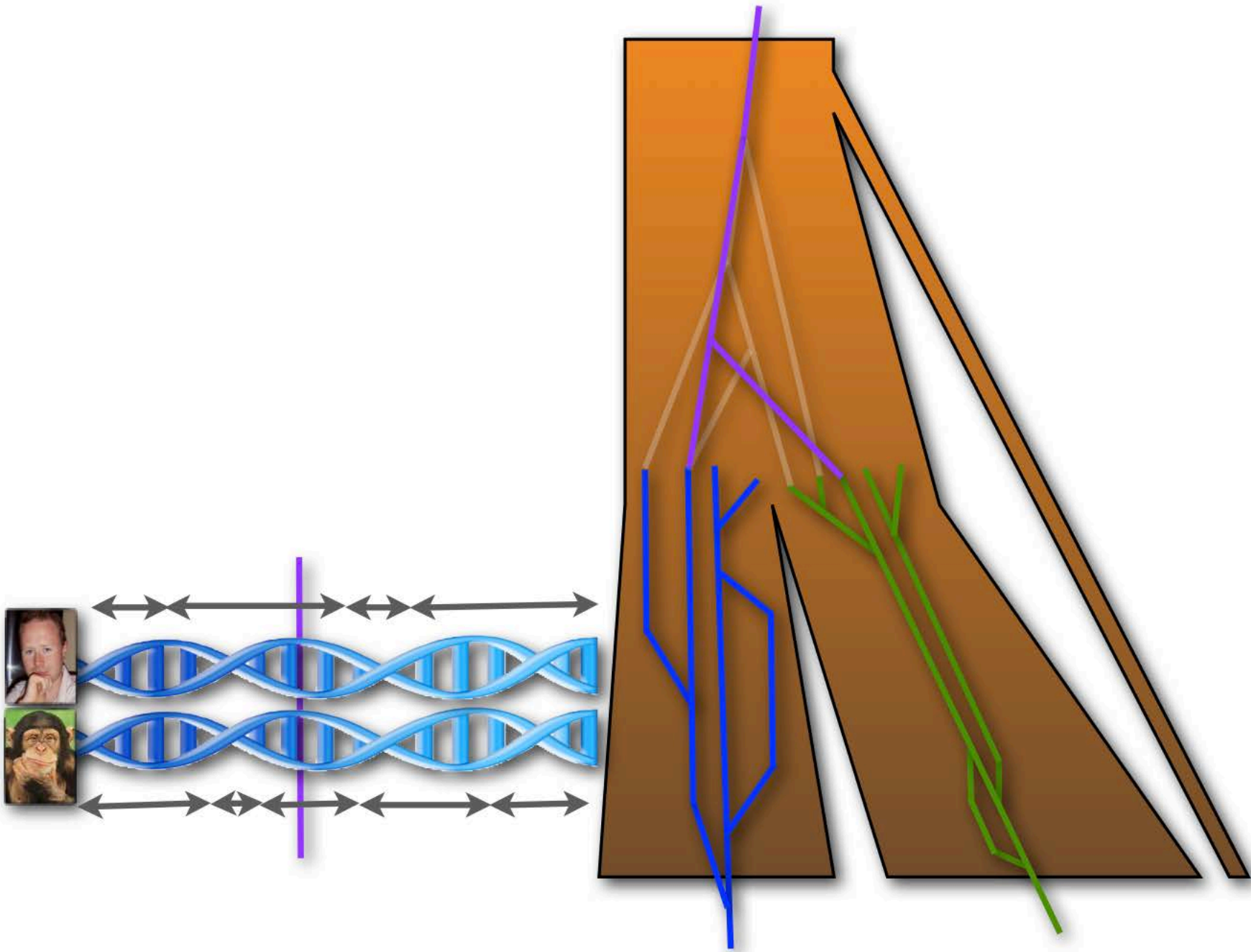


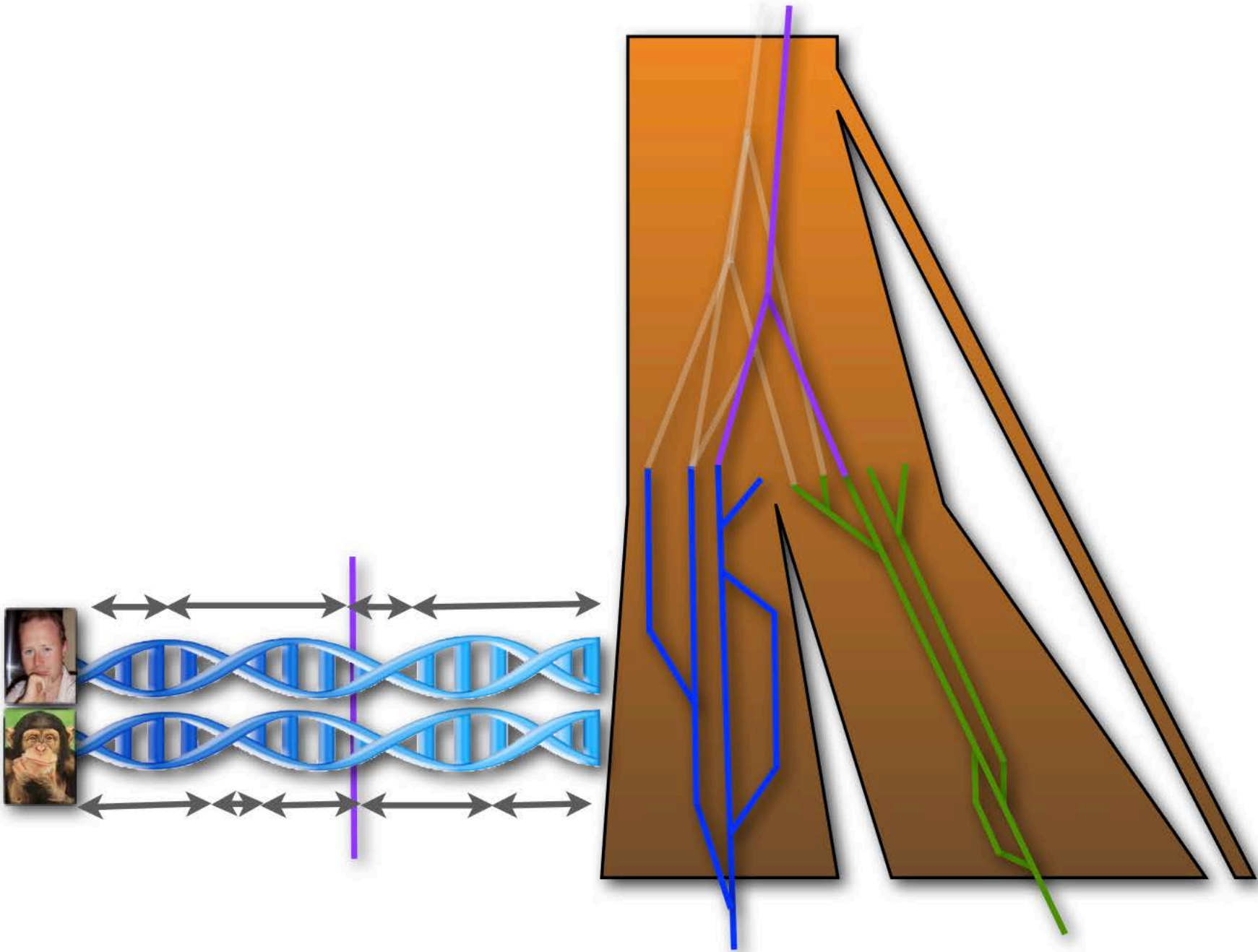


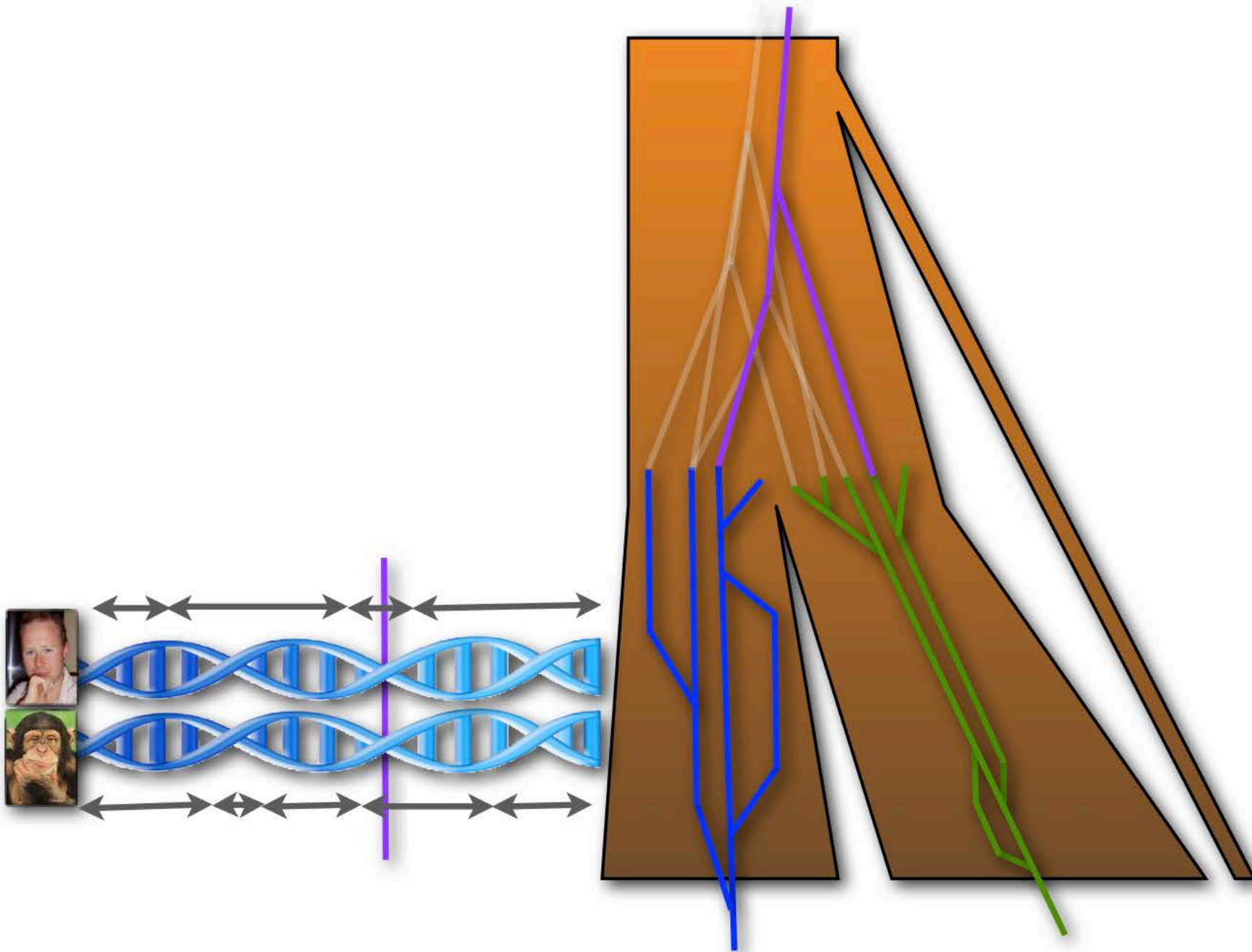


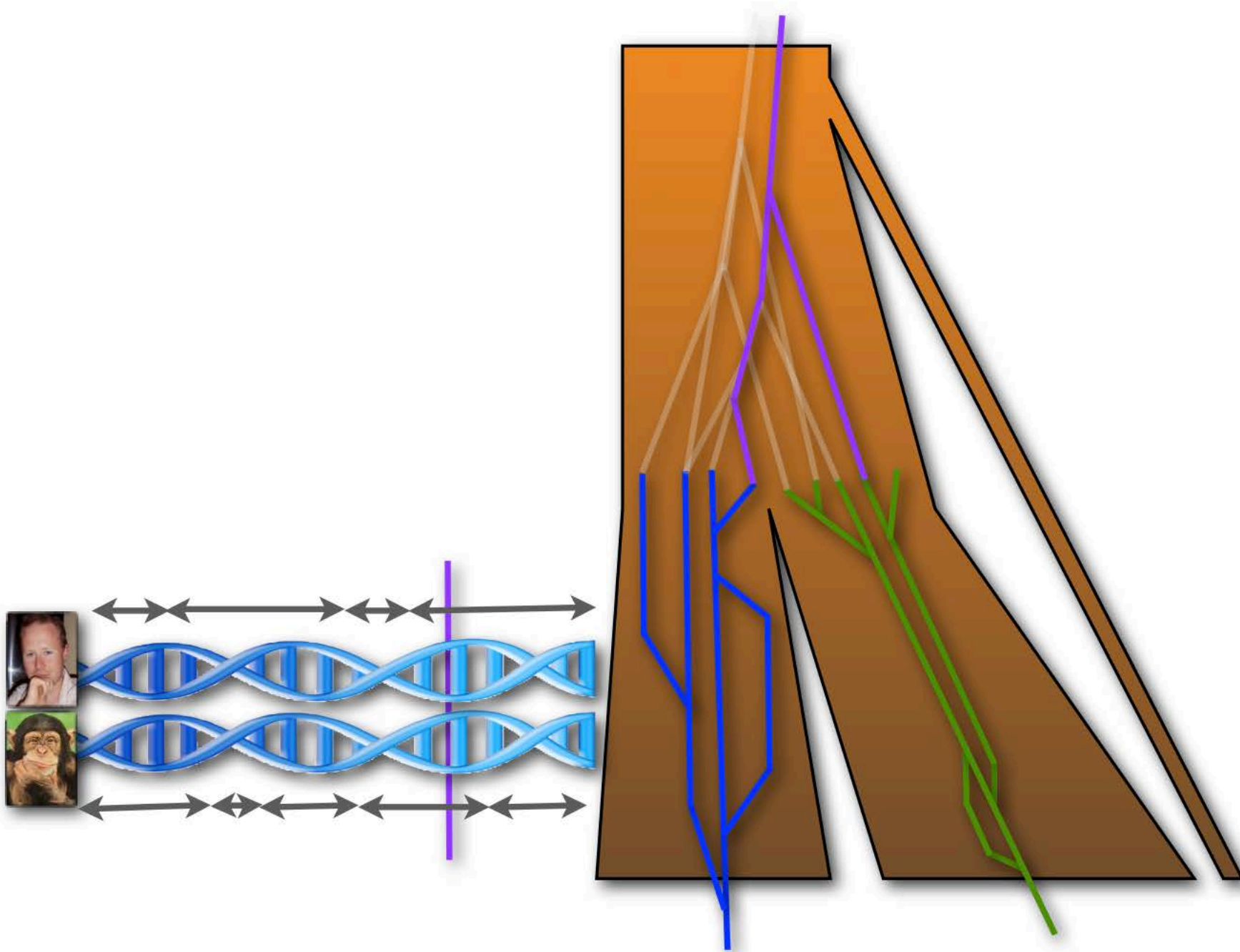


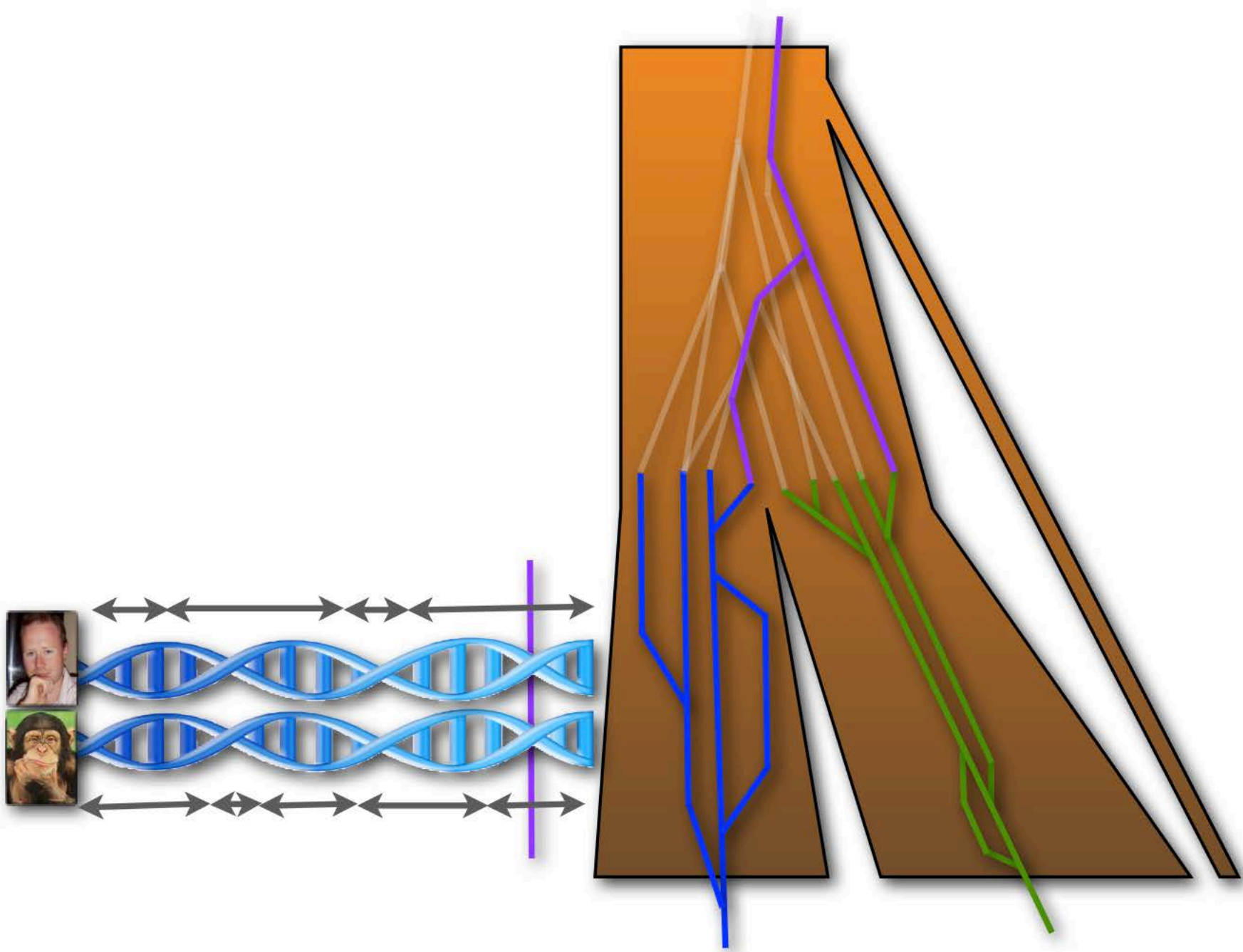


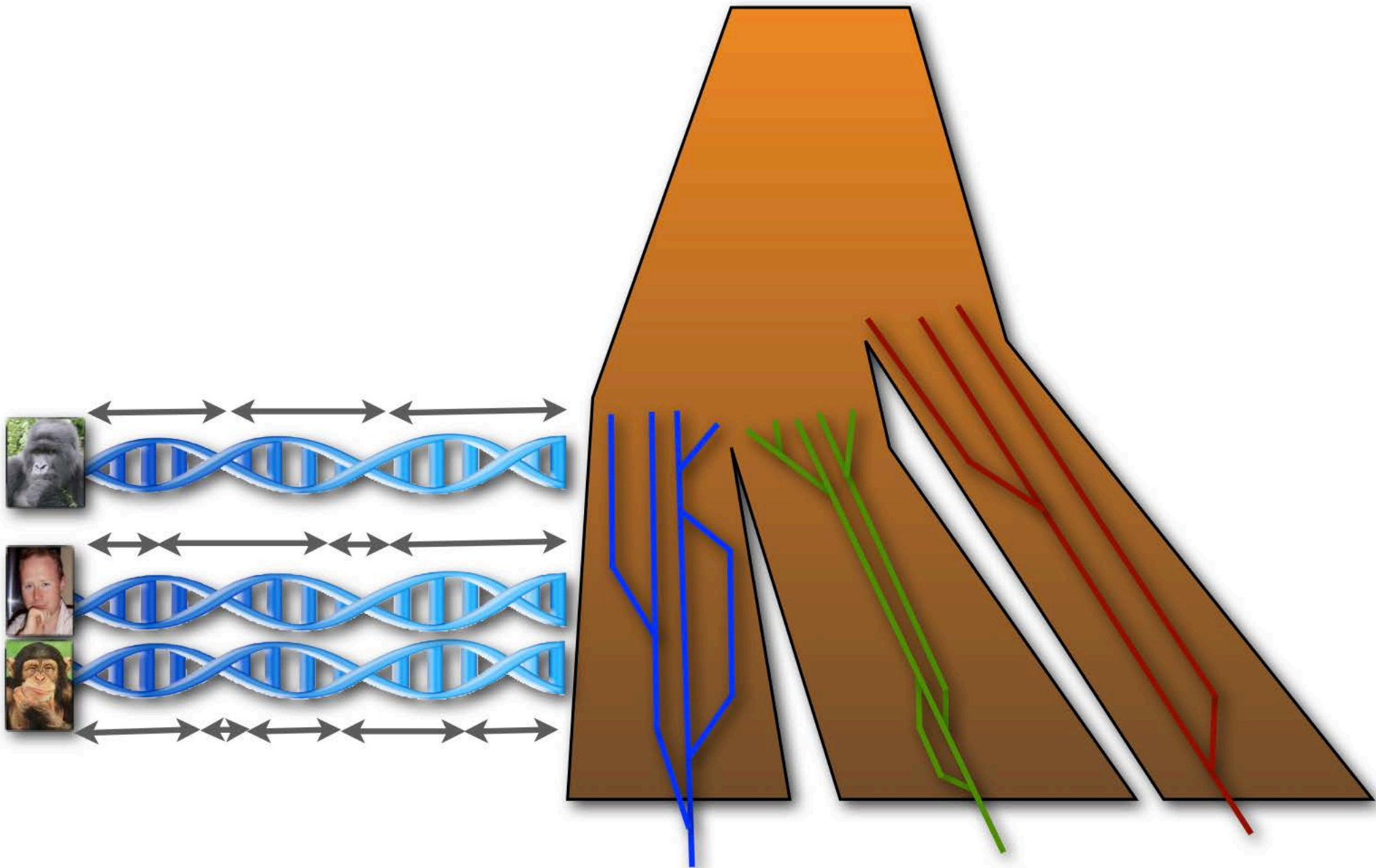


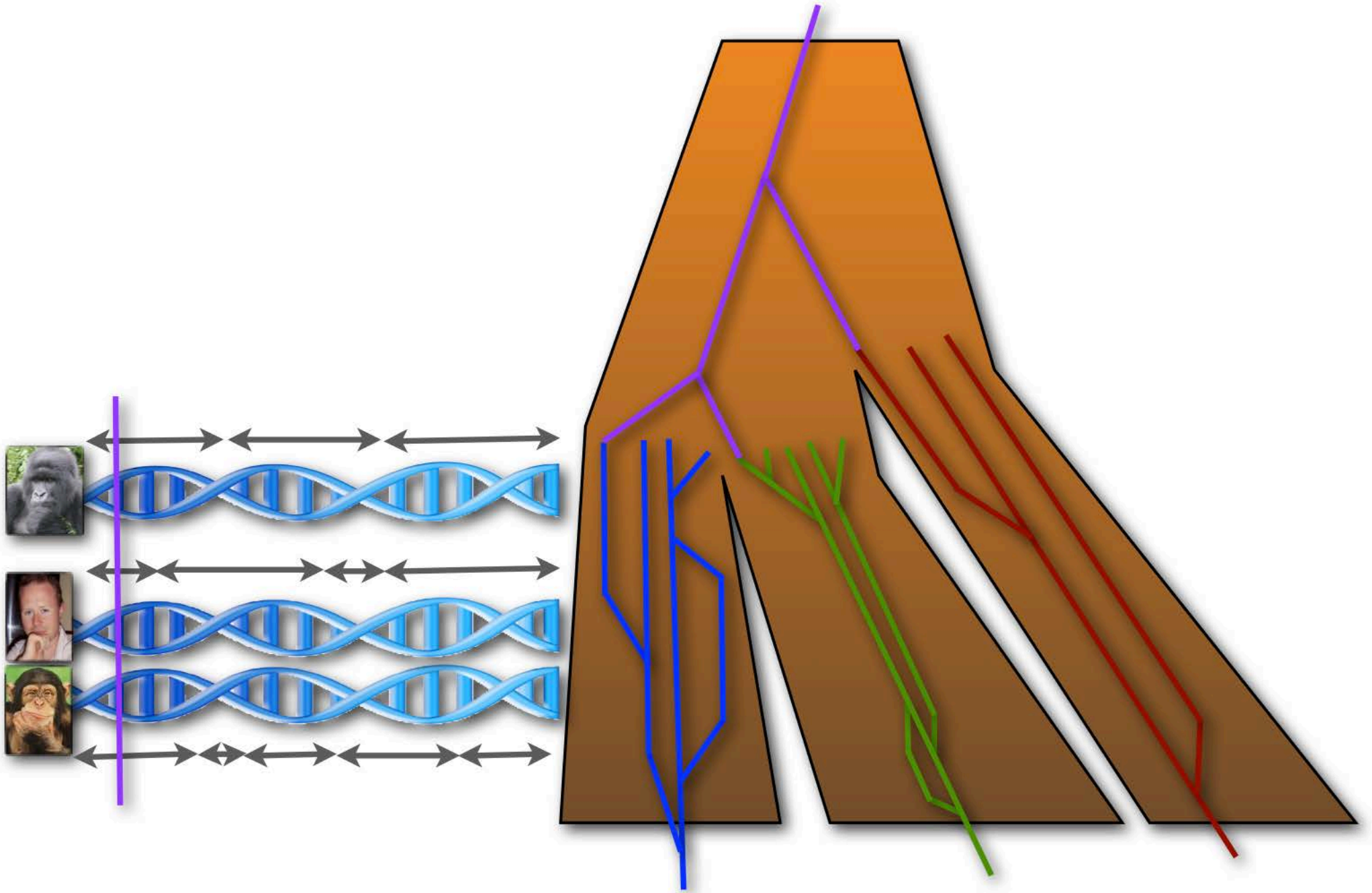


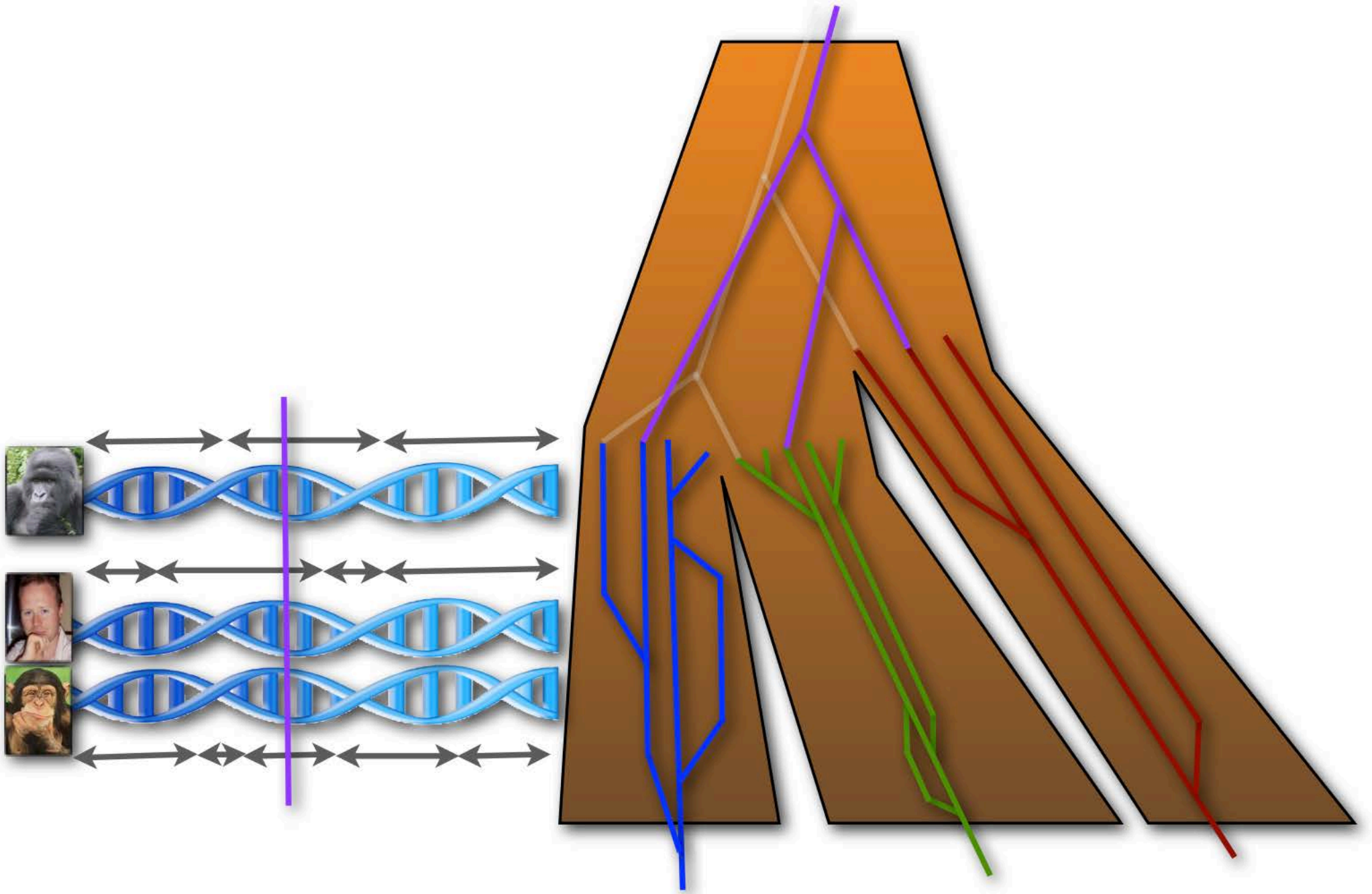












Sequential coalescence models...

Coalescent with recombination

Hudson 1983 - Process running “back in time”

Wiuf & Hein 1999 - Process running “along the sequence”

Coalescent with recombination

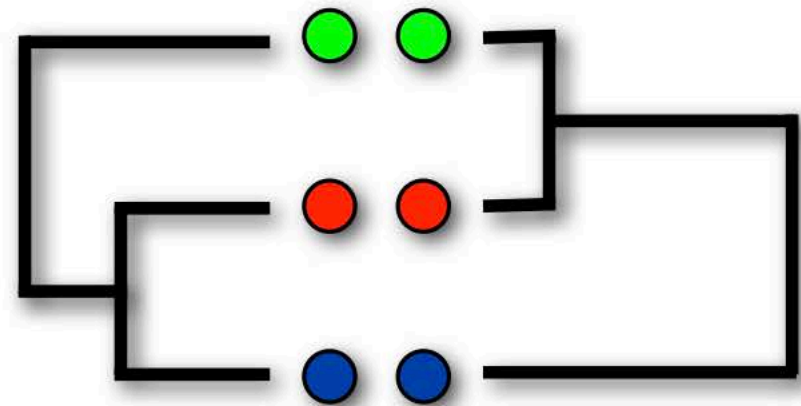
Hudson 1983 - Process running “back in time”

Wiuf & Hein 1999 - Process running “along the sequence”

Sequential Markov
Coalescent

McVean & Cardin 2005
Marjoram & Wall 2006

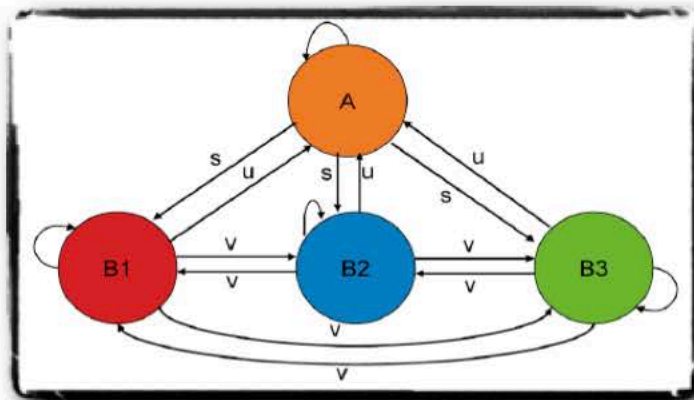
$$P(\mathcal{G}_i | \mathcal{G}_{i-1})$$



Coalescent with recombination

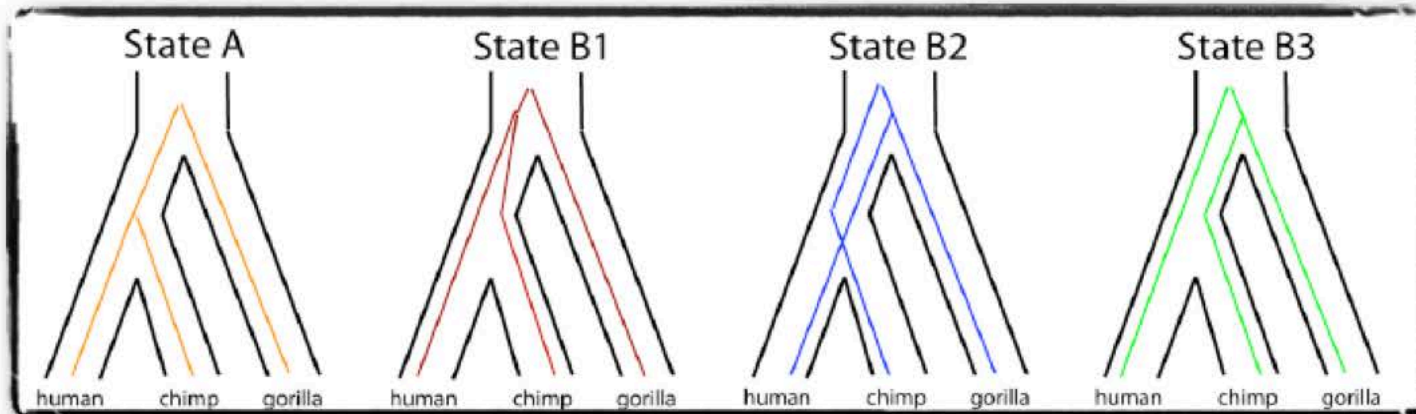
Hudson 1983 - Process running “back in time”

Wiuf & Hein 1999 - Process running “along the sequence”



CoalHMM

Hobolth et al 2007



Coalescent with recombination

Hudson 1983 - Process running “back in time”

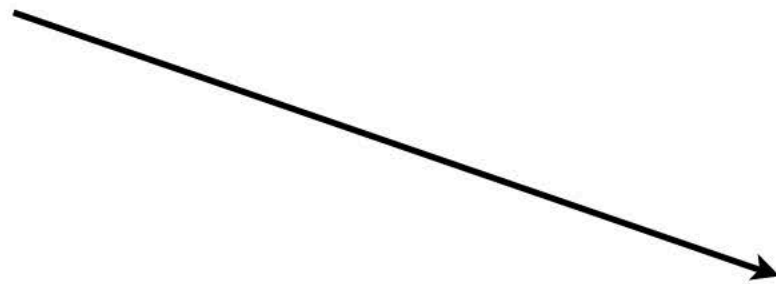
Wiuf & Hein 1999 - Process running “along the sequence”

Sequential Markov
Coalescent

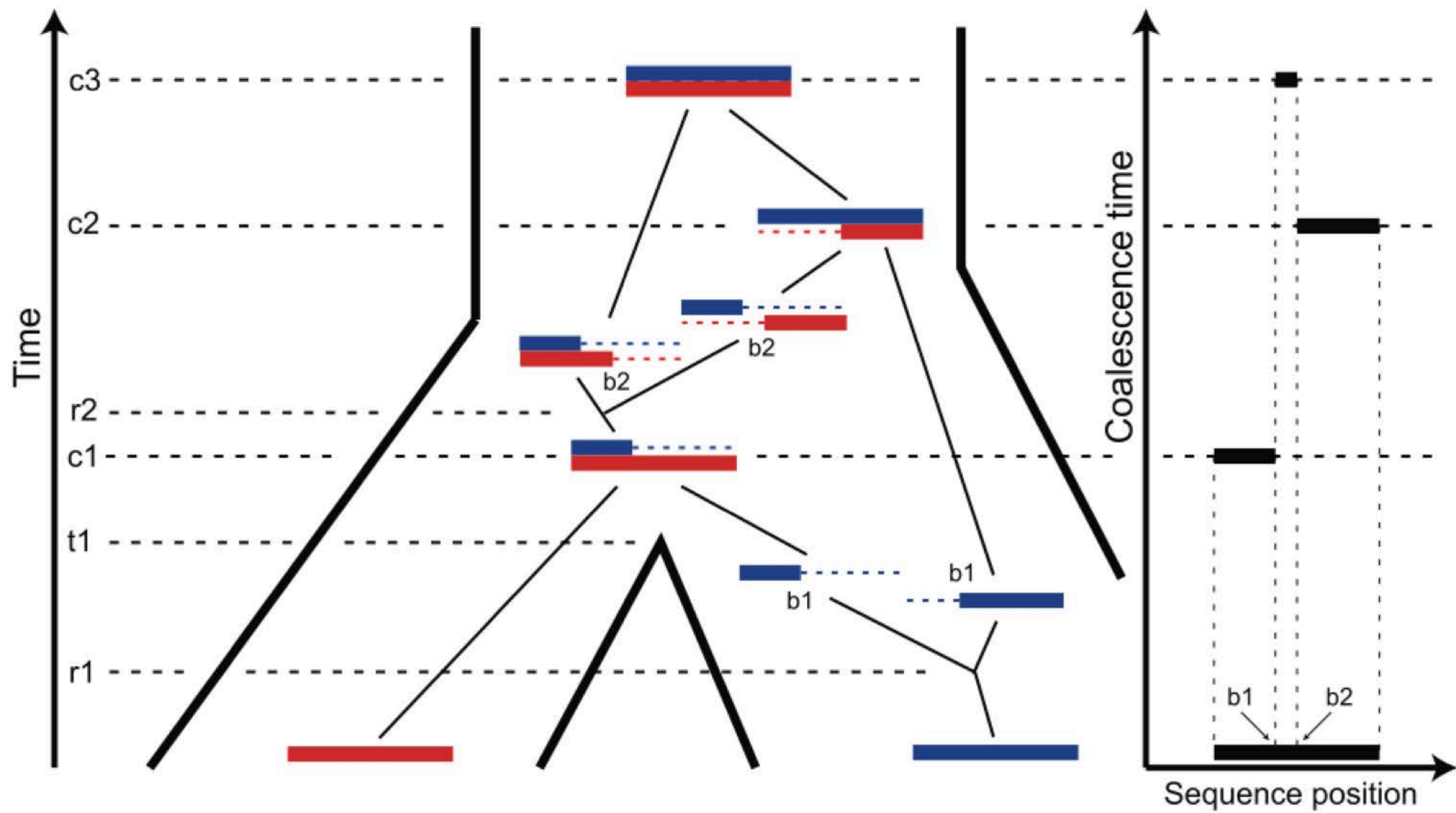
CoalHMM

Hobolth et al 2007

McVean & Cardin 2005
Marjoram & Wall 2006



Dutheil et al 2009

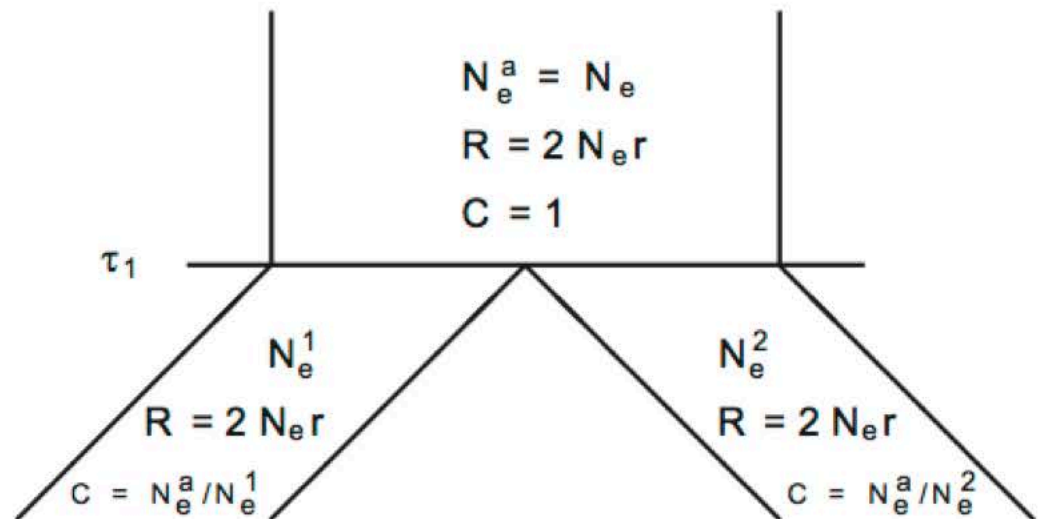


SMC + HMM

Mailund et al 2011

Pairwise SMC/CoalHMM

Inference of split time and
ancestral population size

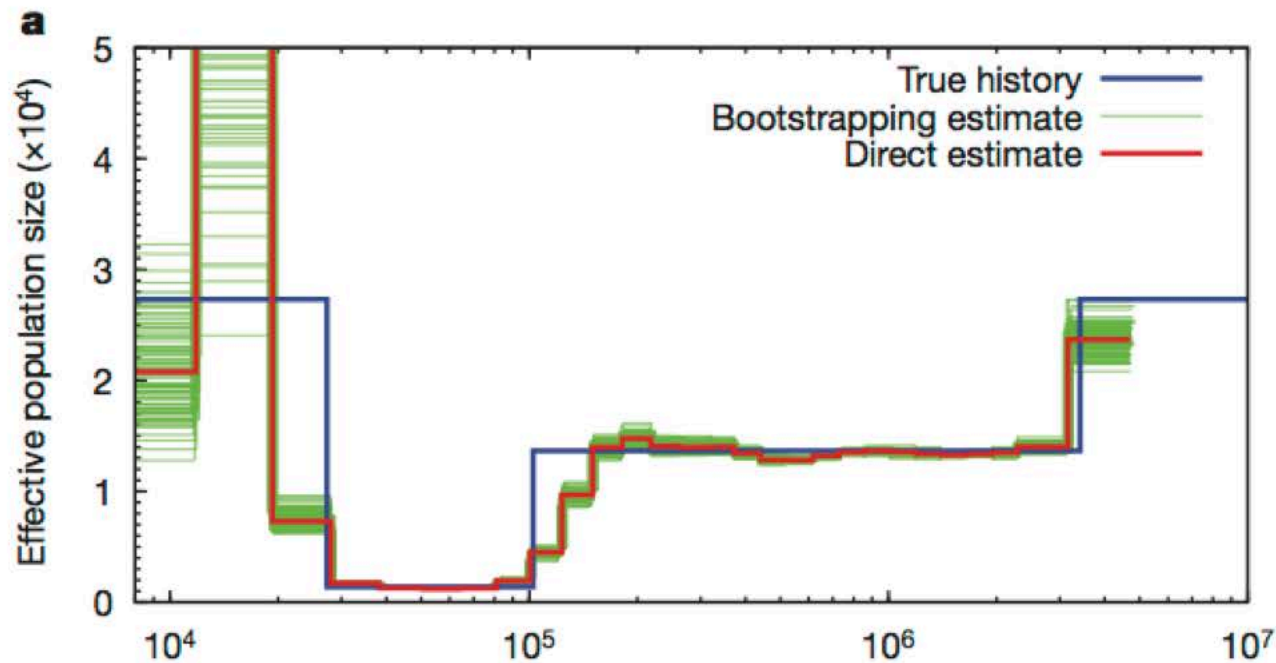


SMC + HMM

Li & Durbin 2011

Pairwise SMC

Inference of effective population size back in time



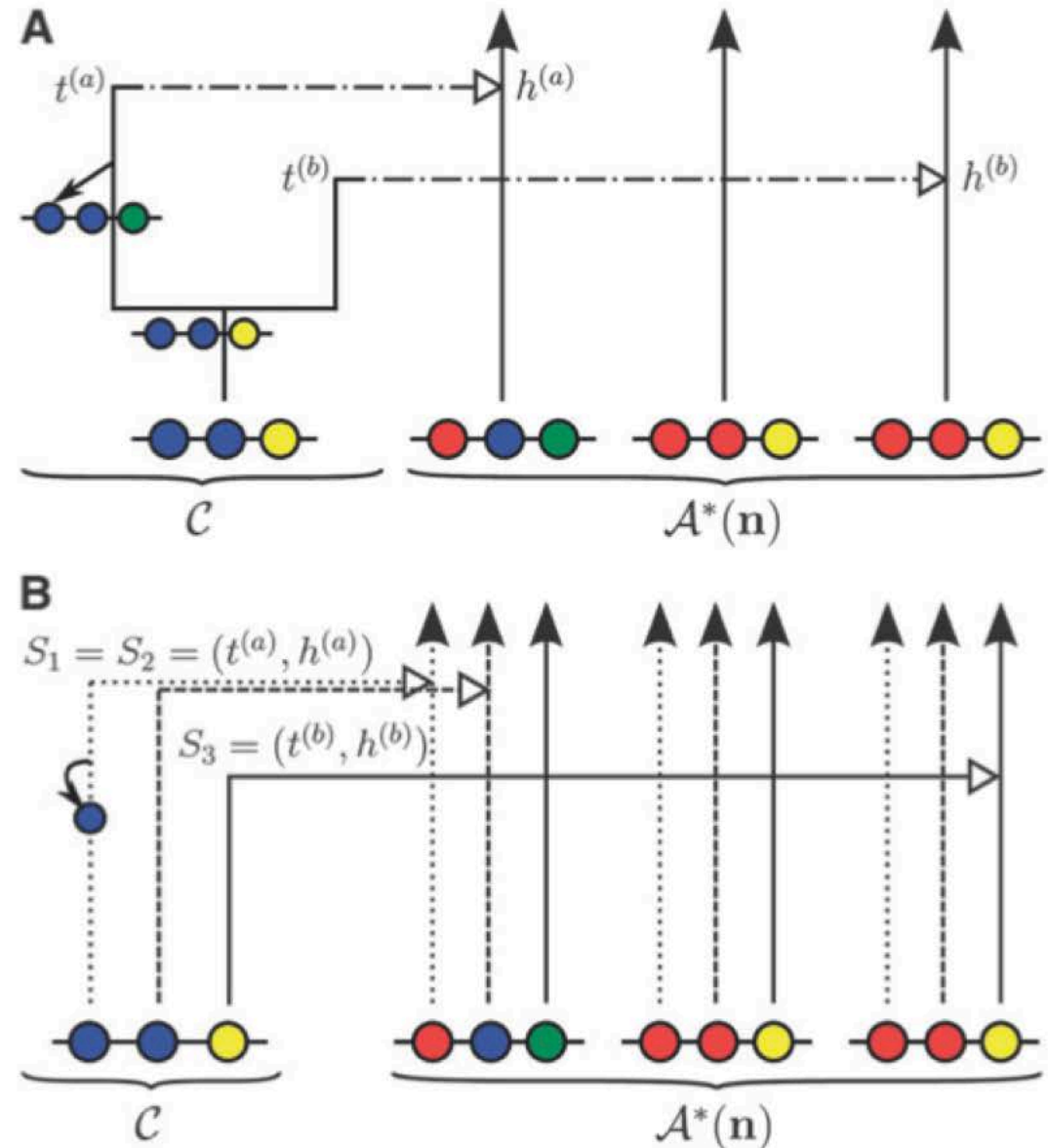
SMC + HMM

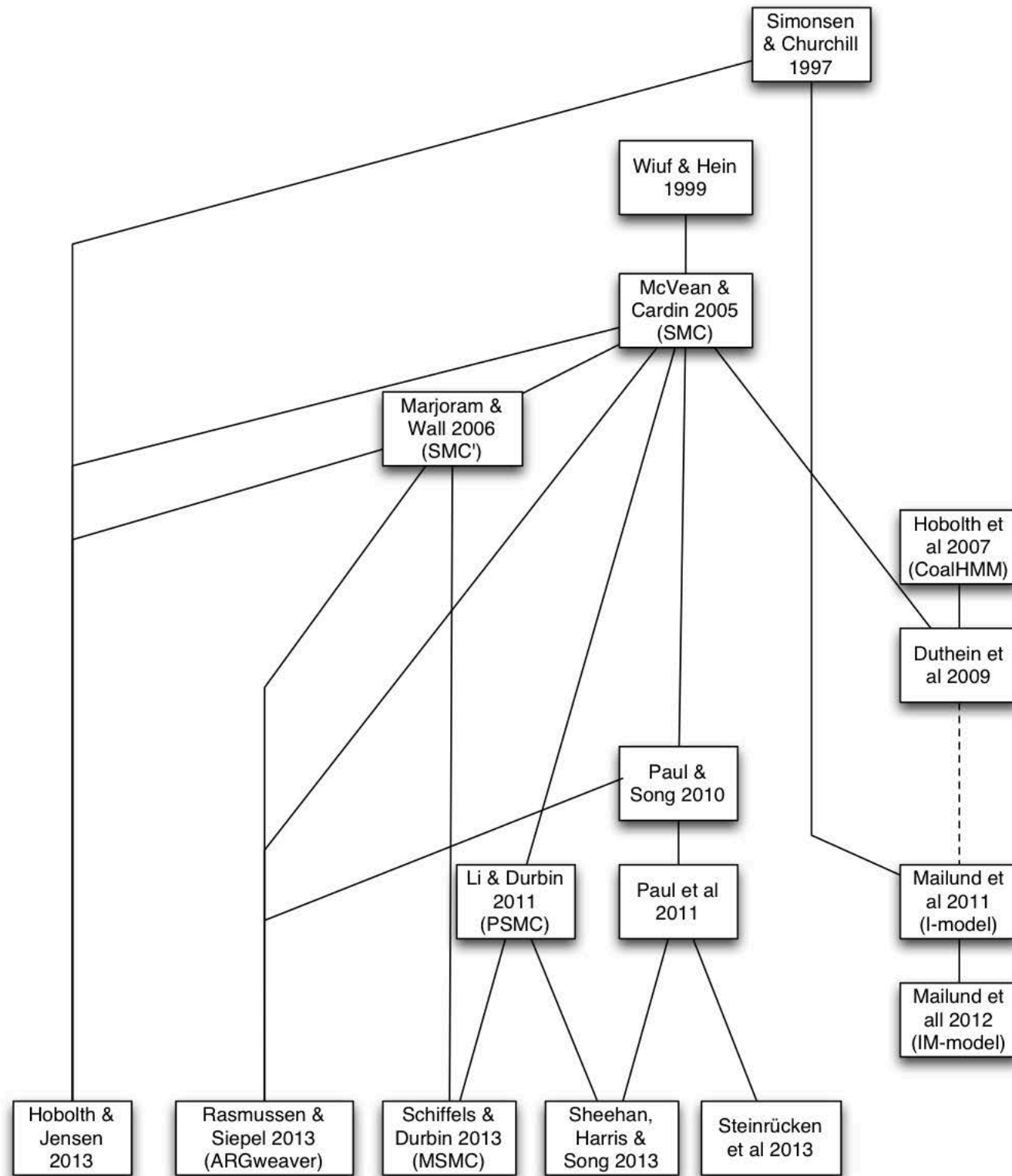
Paul & Song 2010

Paul, Steinrücken and Song 2011

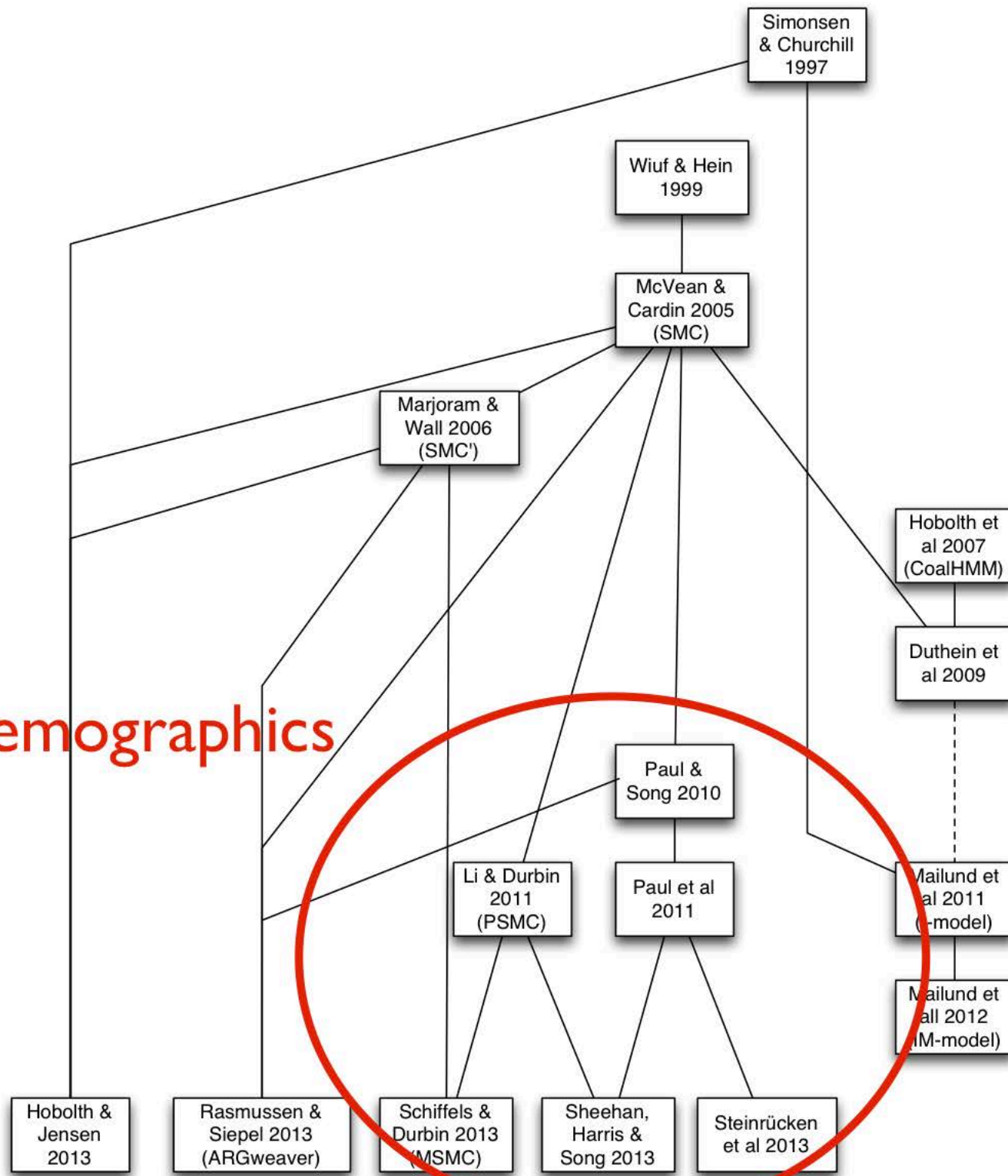
Larger samples but an approximation

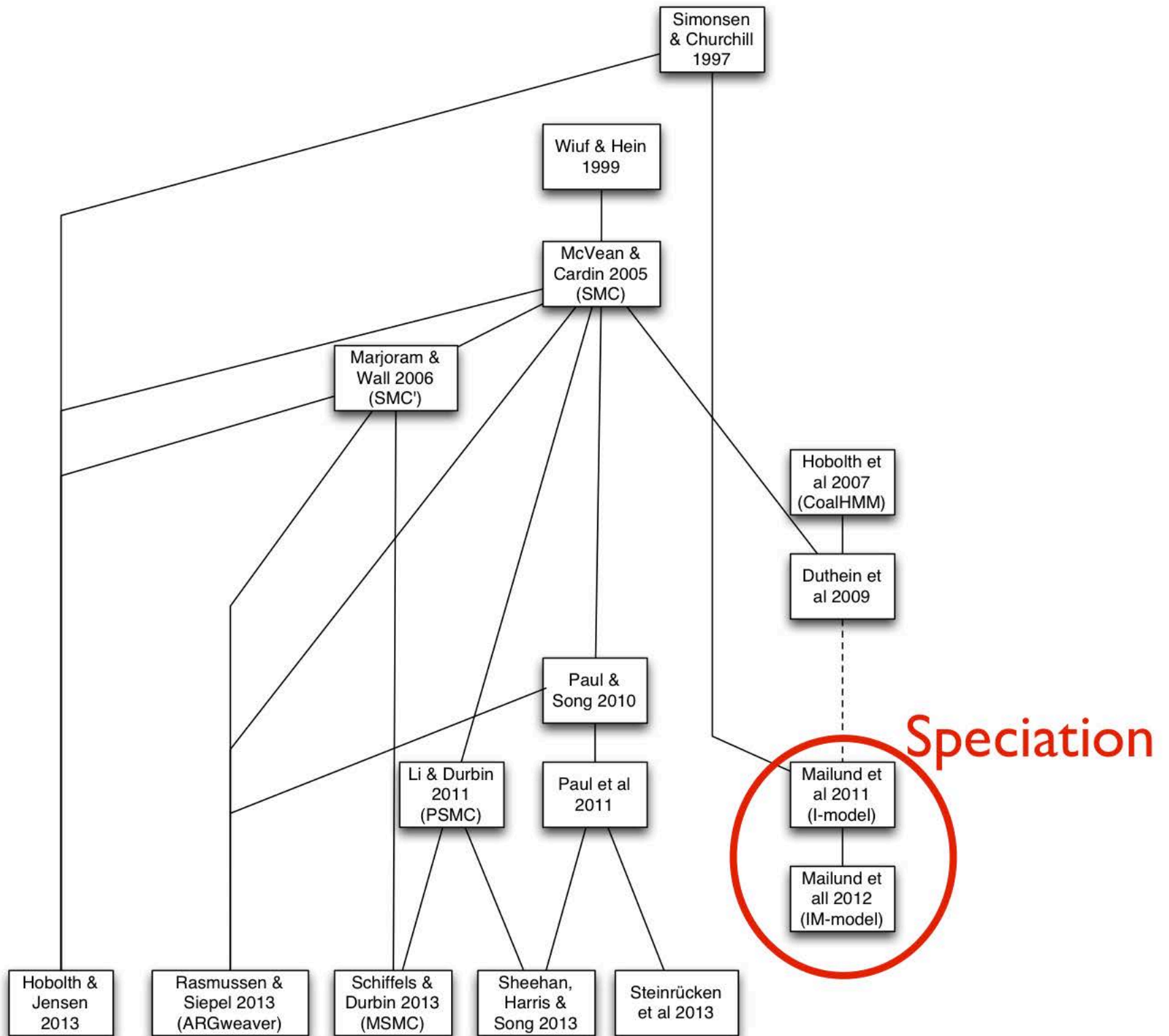
(in many ways still pairwise)



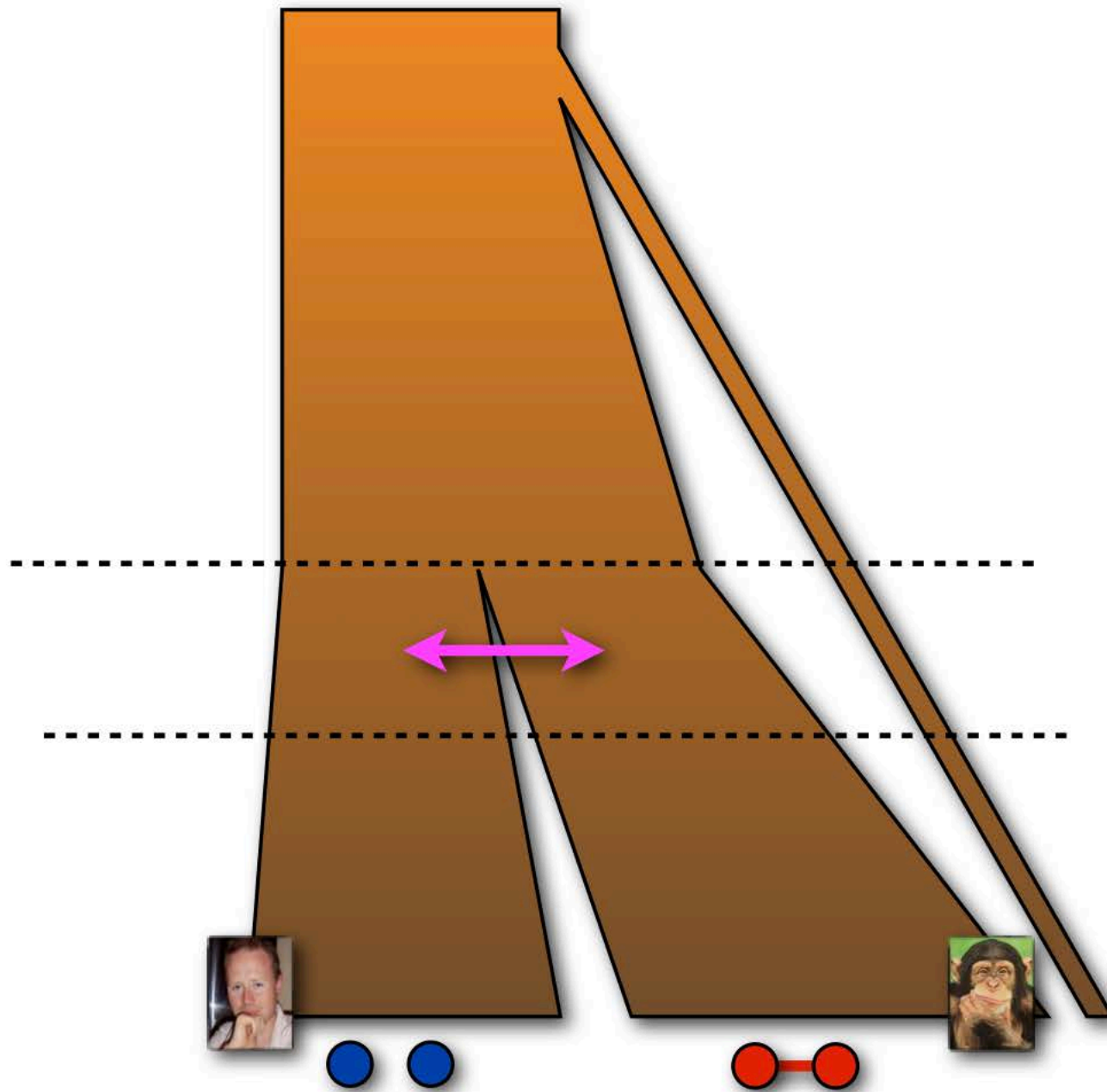


Demographics

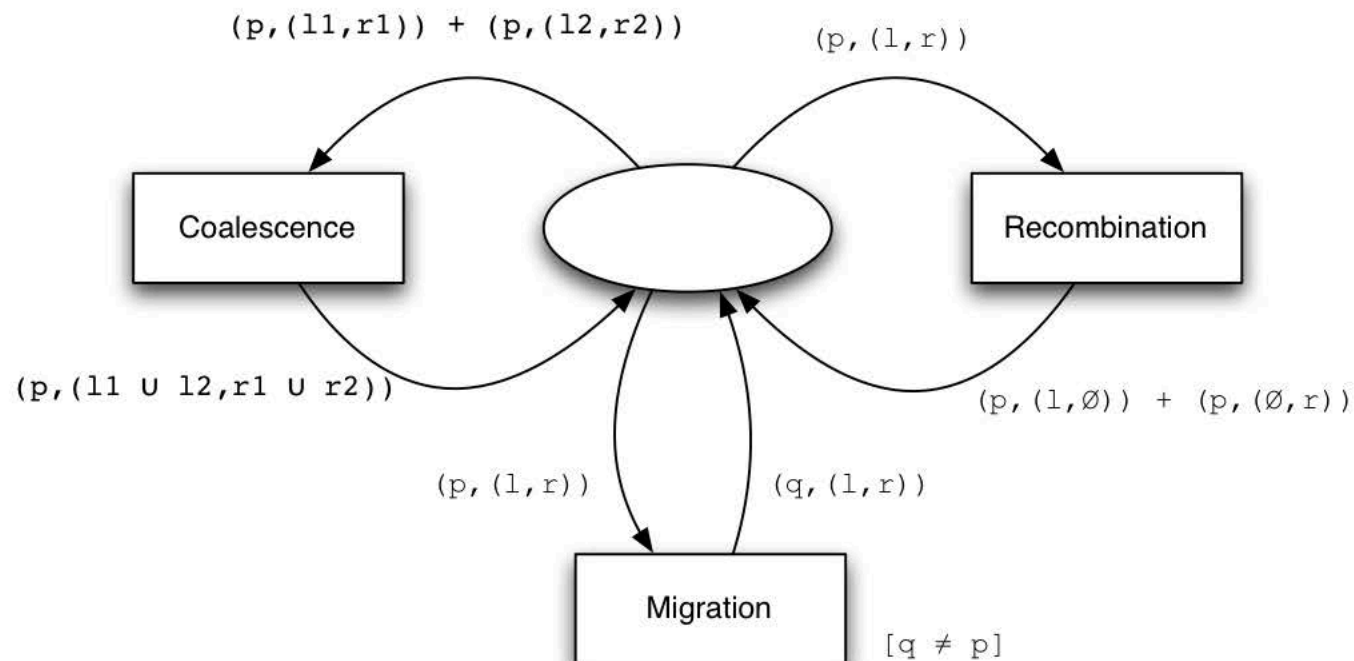


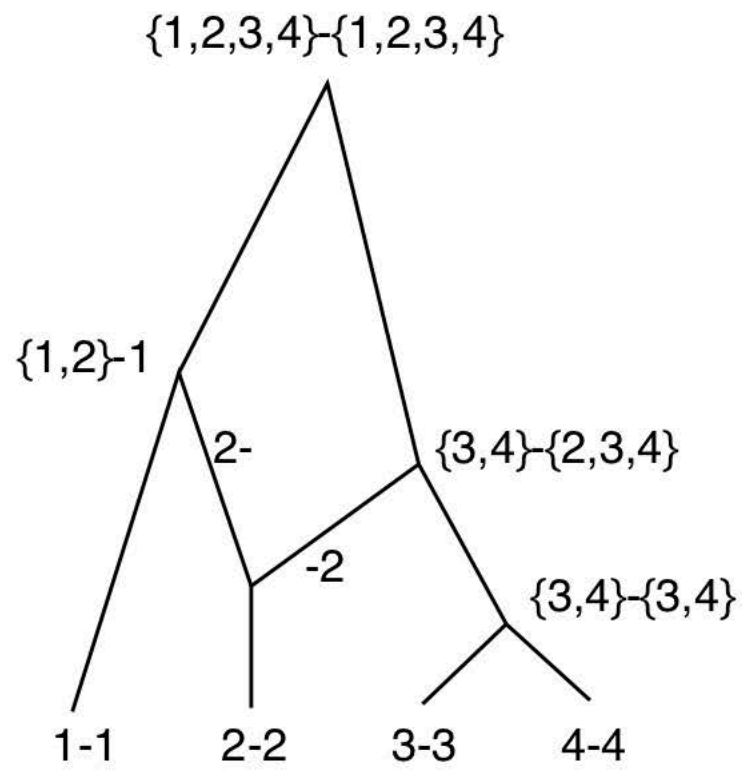


Constructing CoalHMMs



1. **Coalescence:** $\{[p_1, (l_1, r_1)]\} \cup \{[p_2, (l_2, r_2)]\} \cup \mathcal{S} \rightarrow \{[p_1, (l_1 \cup l_2, r_1 \cup r_2)]\} \cup \mathcal{S}$ if $p_1 = p_2$
2. **Recombination:** $\{[p, (l, r)]\} \cup \mathcal{S} \rightarrow \{[p, (l, \emptyset)]\} \cup \{[p, (\emptyset, r)]\} \cup \mathcal{S}$
3. **Migration:** $\{[p_1, (l, r)]\} \cup \mathcal{S} \rightarrow \{[p_2, (l, r)]\} \cup \mathcal{S}$ if $p_1 \neq p_2$.





$$1`(\{1,2,3,4\},\{1,2,3,4\})$$

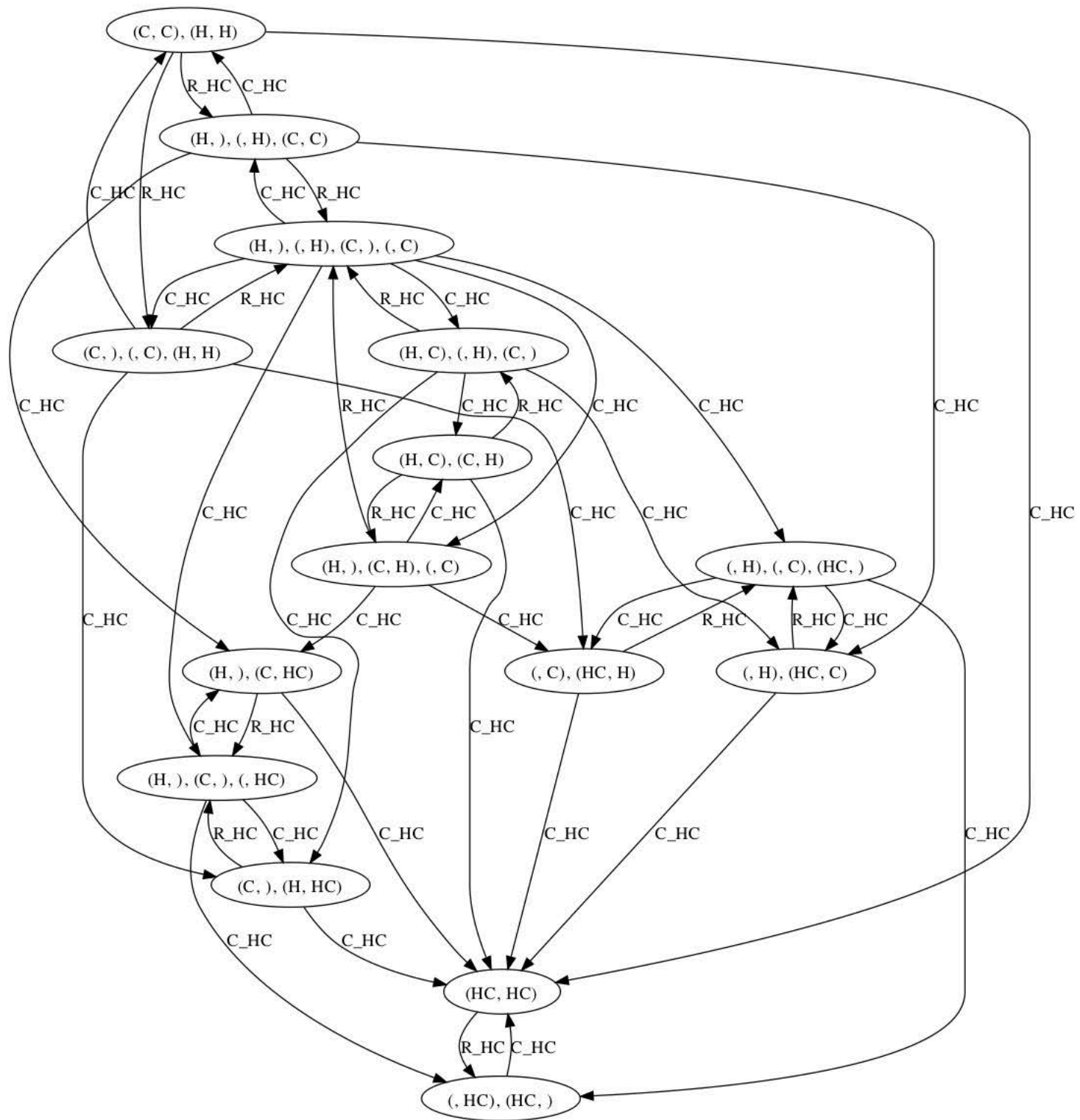
$$1`(\{1,2\},\{1\}) + 1`(\{3,4\},\{2,3,4\})$$

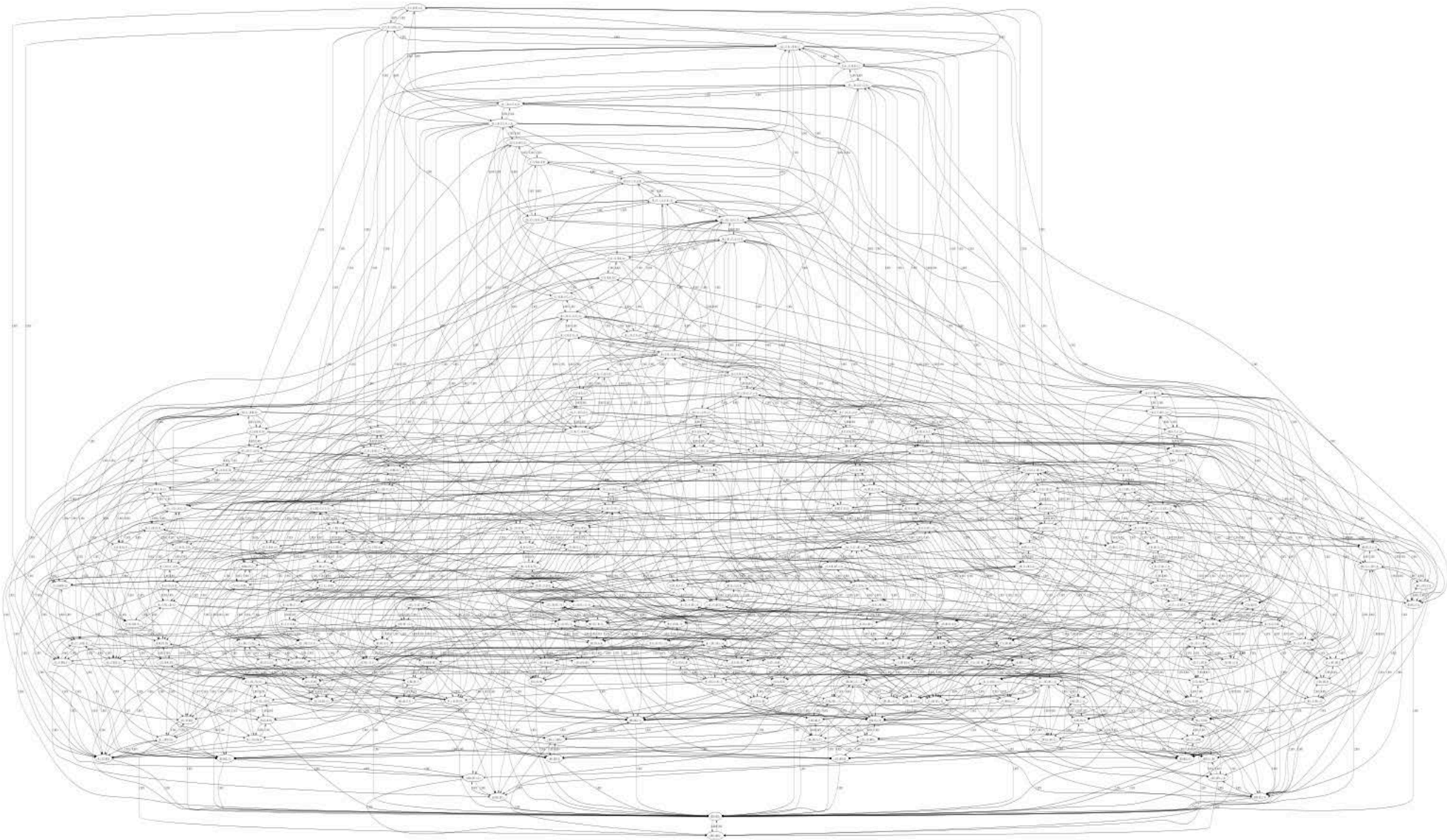
$$1`(\{1\},\{1\}) + 1`(\{2\},\emptyset) + 1`(\{3,4\},\{2,3,4\})$$

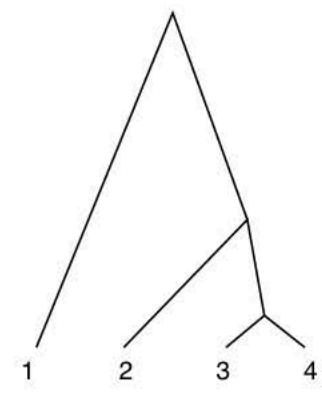
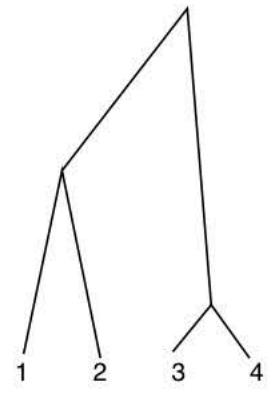
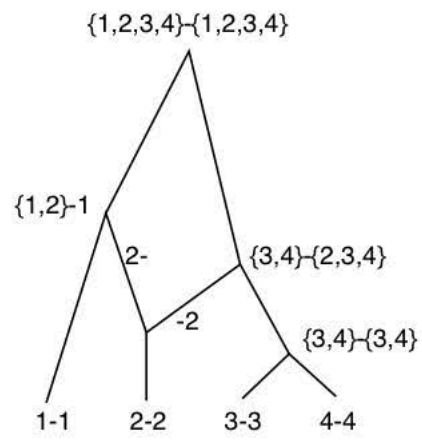
$$1`(\{1\},\{1\}) + 1`(\{2\},\emptyset) + 1`(\emptyset,\{2\}) + 1`(\{3,4\},\{3,4\})$$

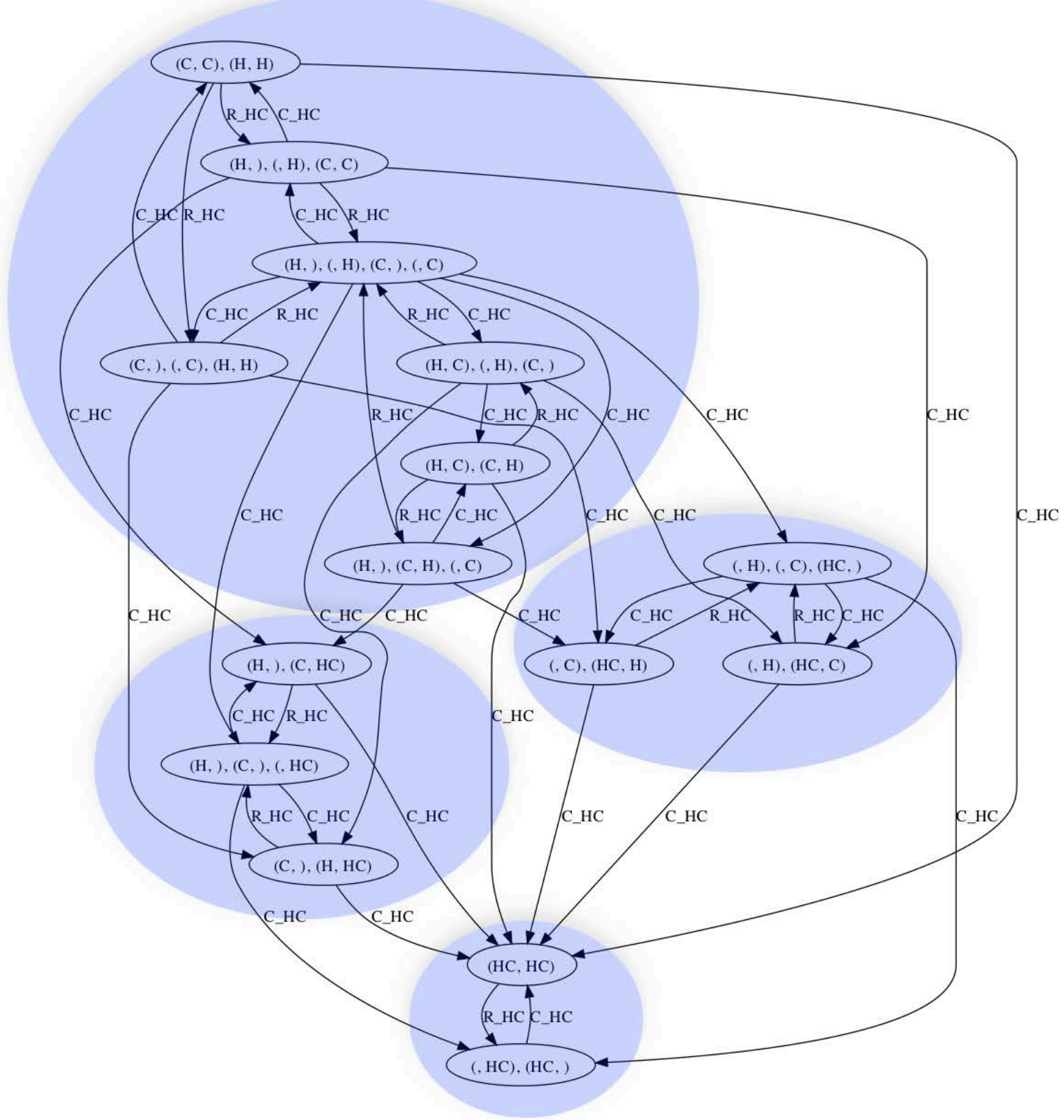
$$1`(\{1\},\{1\}) + 1`(\{2\},\{2\}) + 1`(\{3,4\},\{3,4\})$$

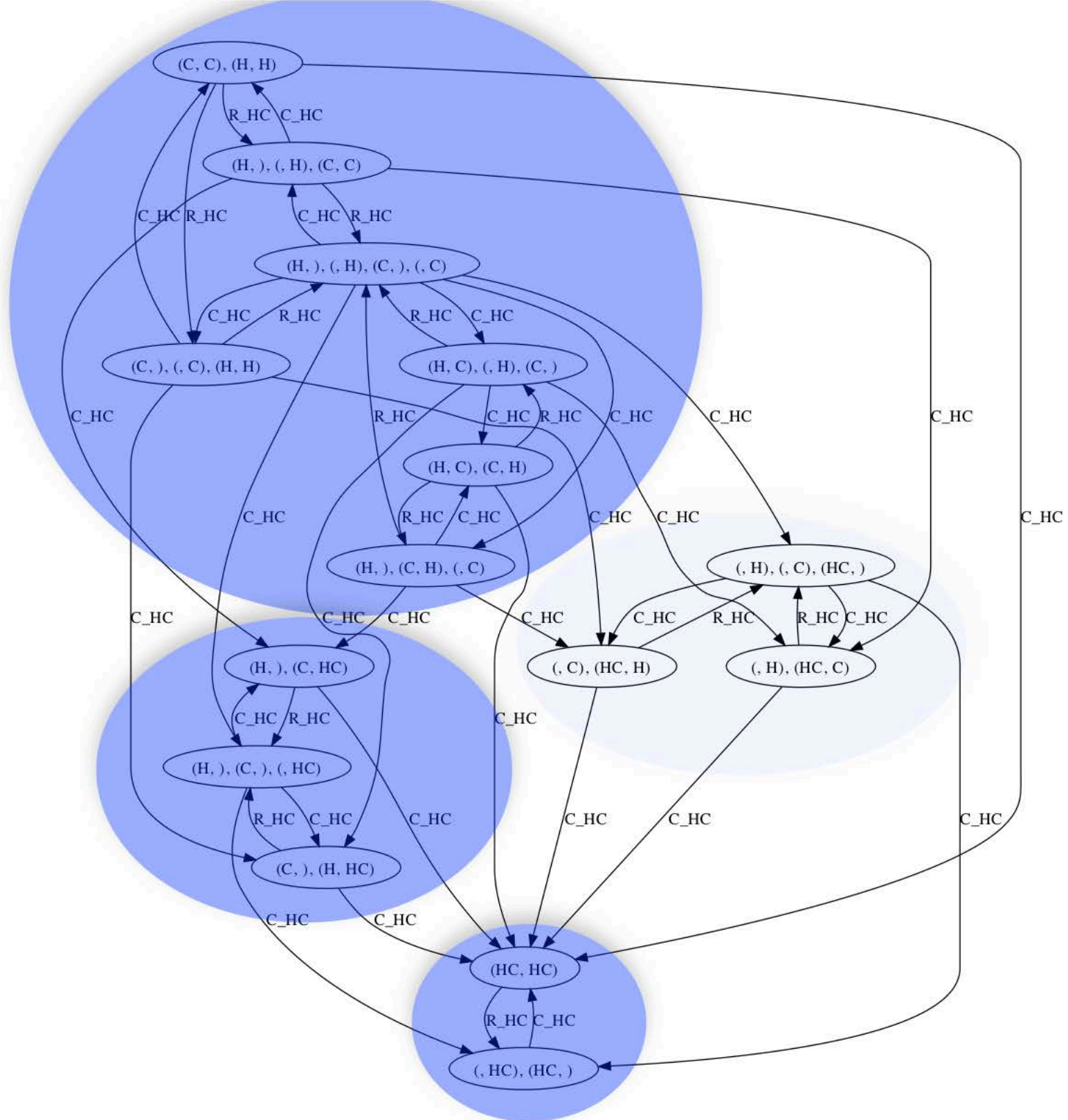
$$1`(\{1\},\{1\}) + 1`(\{2\},\{2\}) + 1`(\{3\},\{3\}) + 1`(\{4\},\{4\})$$

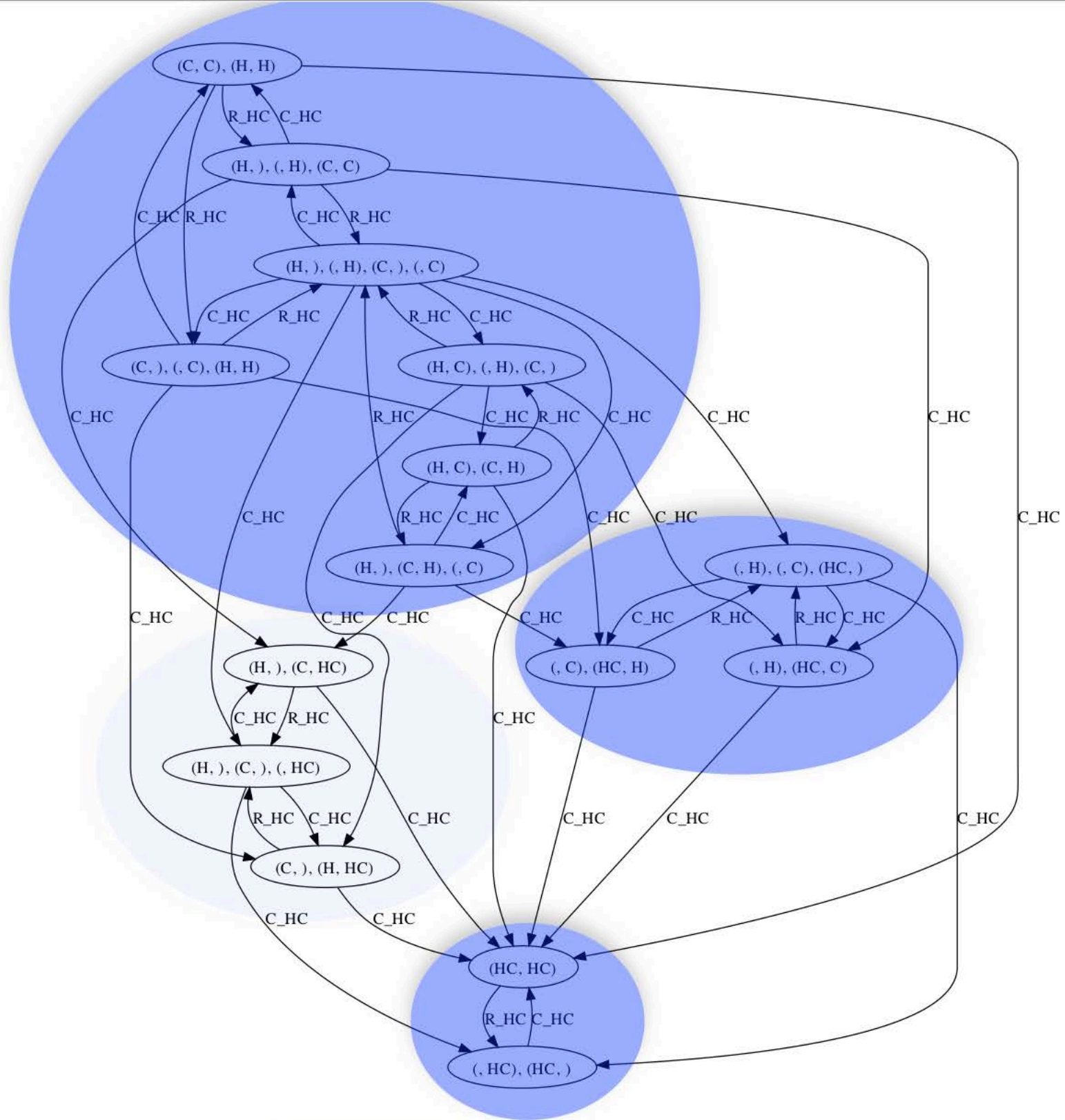


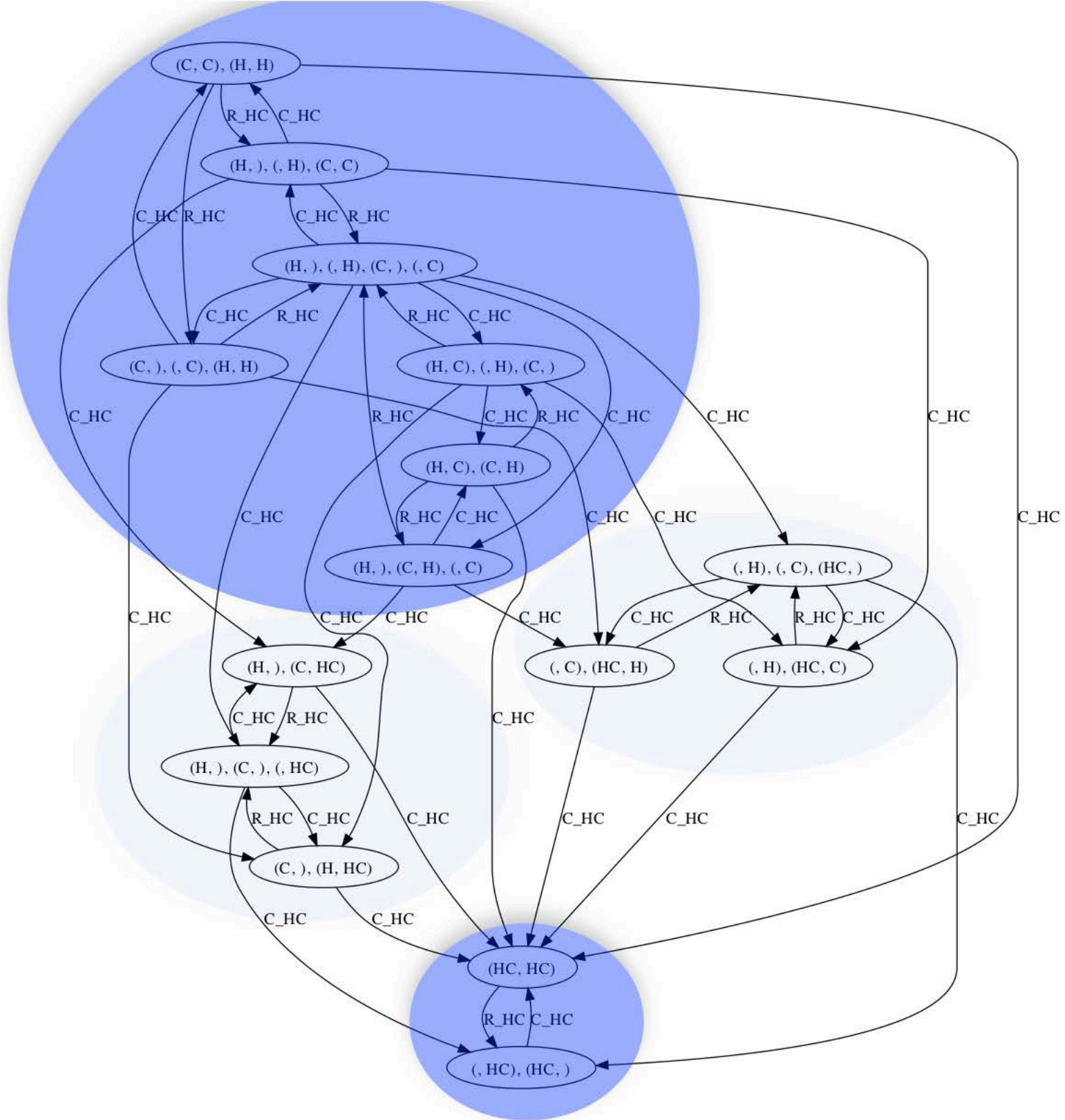


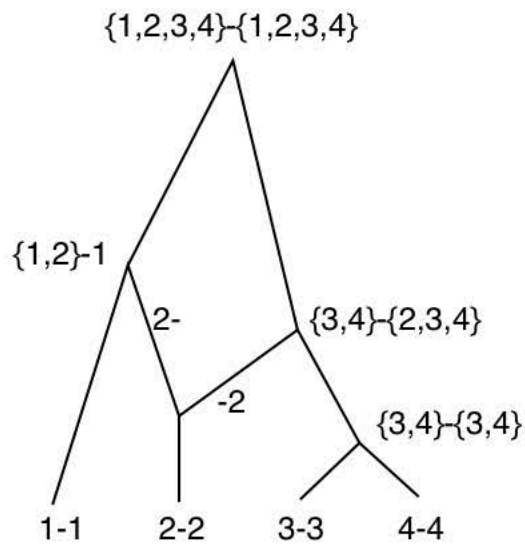












$$1^{\{1,2,3,4\},\{1,2,3,4\}}$$

$$1^{\{1,2\},\{1\}} + 1^{\{3,4\},\{2,3,4\}}$$

$$1^{\{1\},\{1\}} + 1^{\{2\},\emptyset} + 1^{\{3,4\},\{2,3,4\}}$$

$$1^{\{1\},\{1\}} + 1^{\{2\},\emptyset} + 1^{\emptyset,\{2\}} + 1^{\{3,4\},\{3,4\}}$$

$$1^{\{1\},\{1\}} + 1^{\{2\},\{2\}} + 1^{\{3,4\},\{3,4\}}$$

$$1^{\{1\},\{1\}} + 1^{\{2\},\{2\}} + 1^{\{3\},\{3\}} + 1^{\{4\},\{4\}}$$

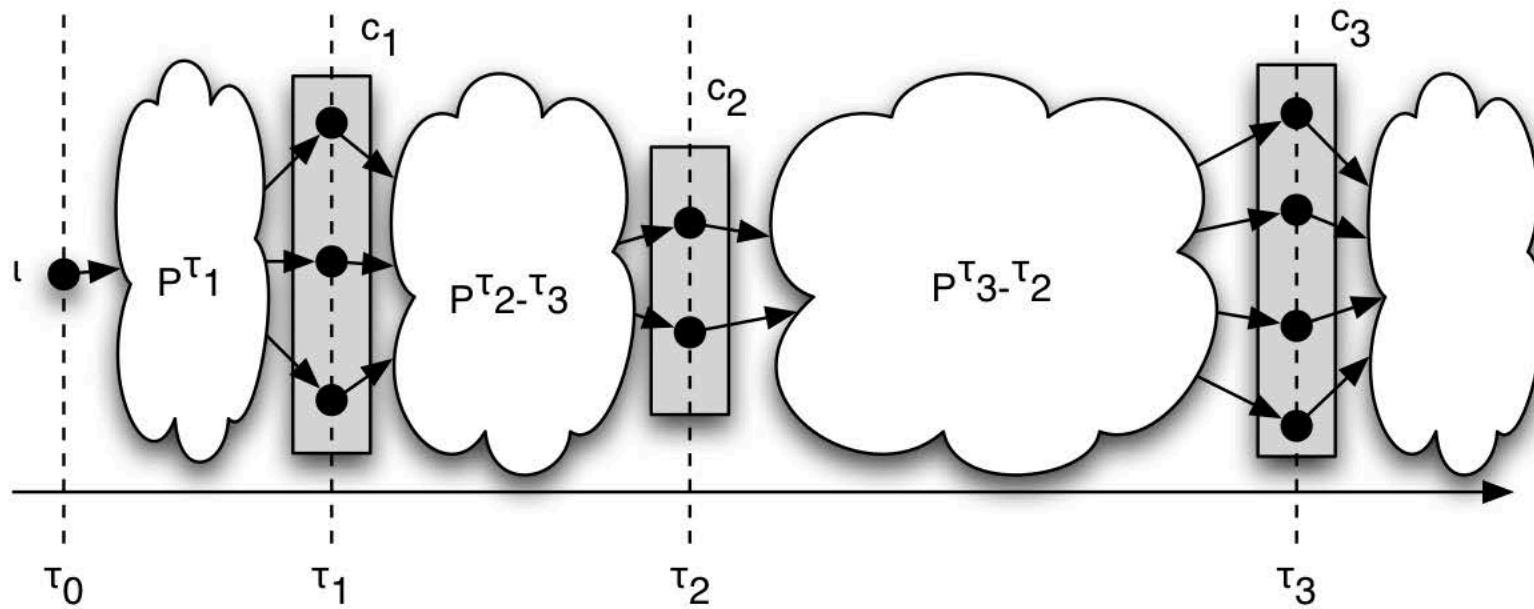


{1,2},{3,4} and {1},{2,3,4}

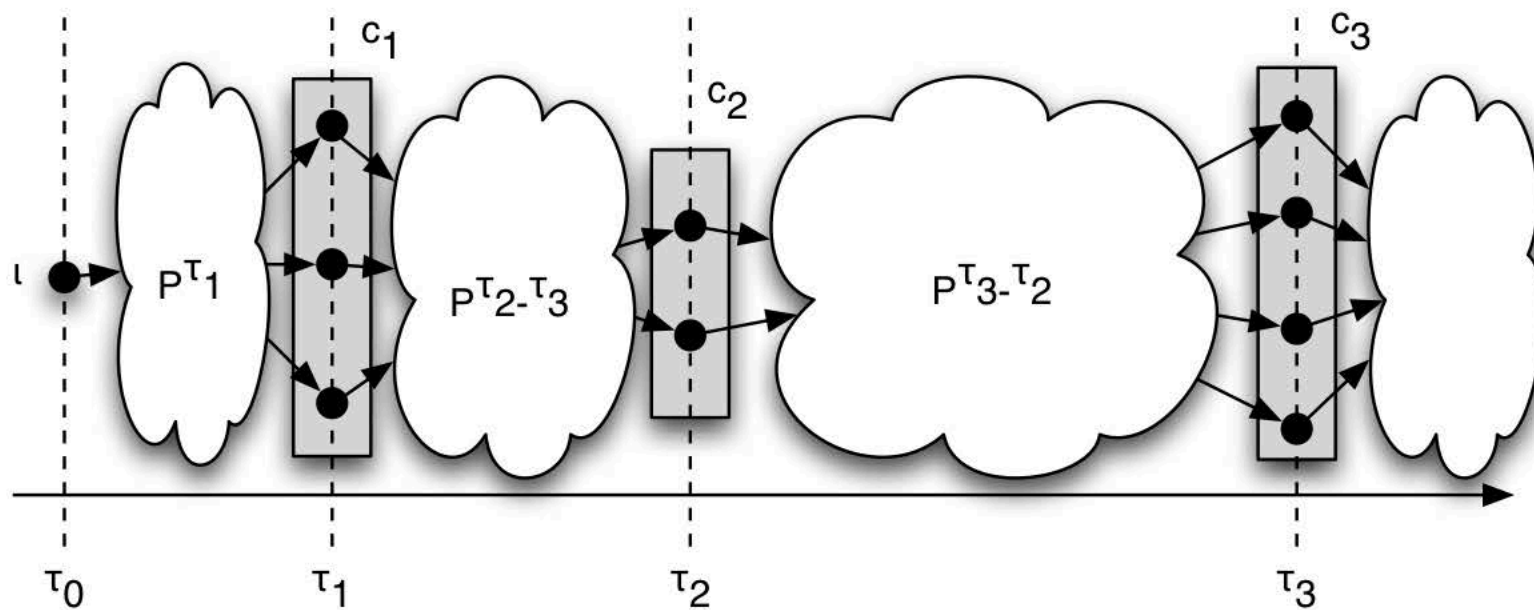
{1},{2},{3,4} and {1},{2,3,4}

{1},{2},{3,4} and {1},{2},{3,4}

{1},{2},{3},{4} and {1},{2},{3},{4}

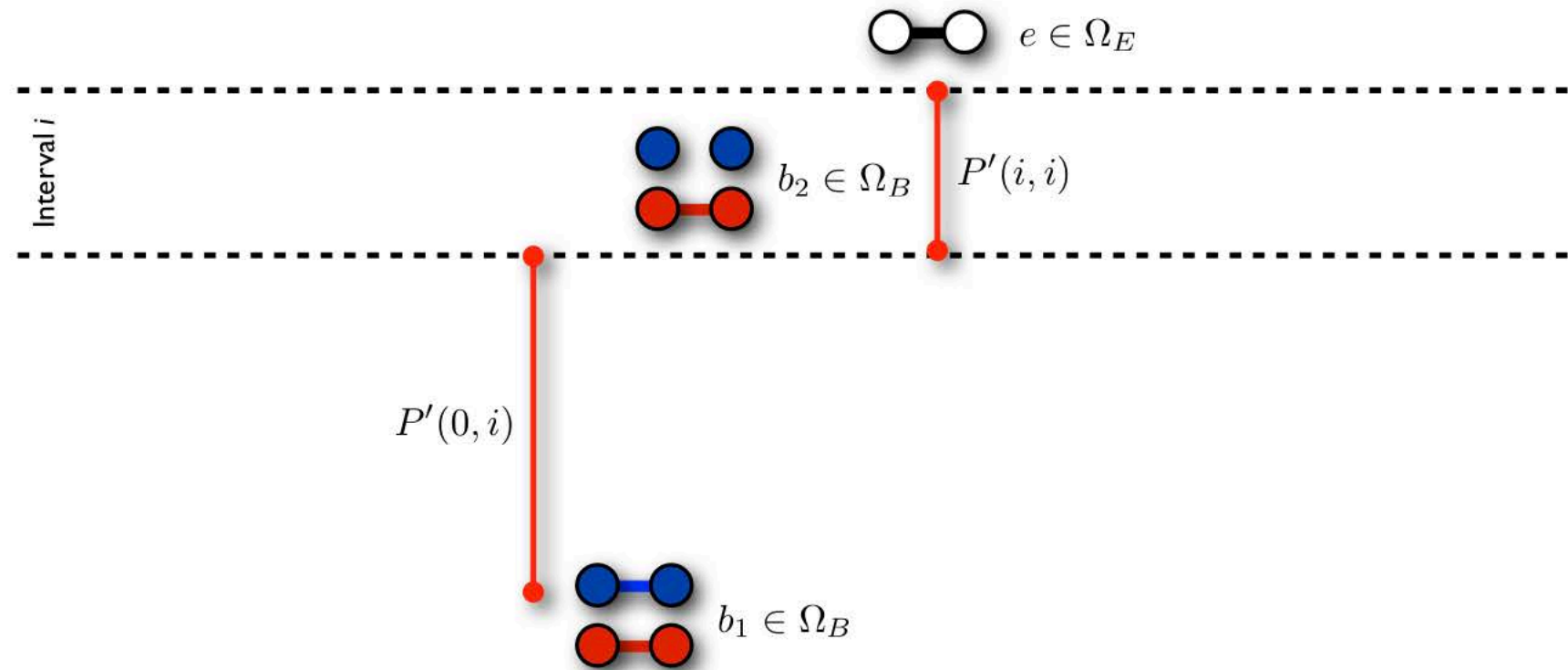


$$\sum_{s_1 \in c_{i_1}} \sum_{s_2 \in c_{i_2}} \dots \sum_{s_{n-1} \in c_{i_{n-1}}} P_{l, s_1}^{\tau_1} \cdot P_{s_{i_1}, s_{i_2}}^{\tau_2 - \tau_1} \dots P_{s_{n-2}, s_{n-1}}^{\tau_{n-1} - \tau_{n-2}}$$

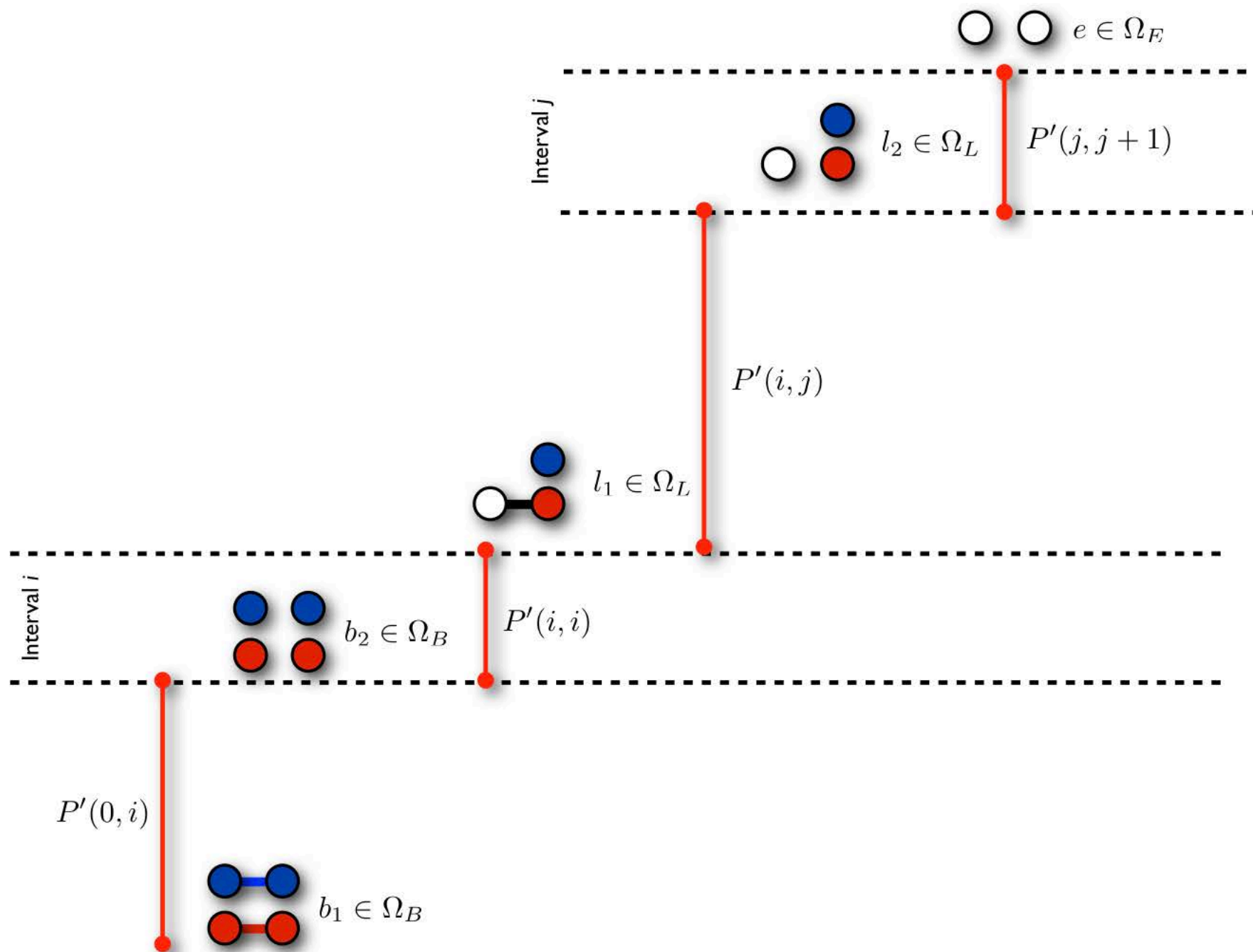


$$\sum_{s_1 \in c_{i_1}} P_{l, s_1}^{\tau_1} \left(\sum_{s_2 \in c_{i_2}} P_{s_{i_1}, s_{i_2}}^{\tau_2 - \tau_1} \left(\dots \left(\sum_{s_{n-1} \in c_{i_{n-1}}} P_{s_{n-2}, s_{n-1}}^{\tau_{n-1} - \tau_{n-2}} \right) \dots \right) \right)$$

$$\Pr(L \in i, R \in i) = \sum_{b_1 \in \Omega_B} \sum_{b_2 \in \Omega_B} \sum_{e \in \Omega_E} P'(0, i)_{b_1, b_2} P'(i, i+1)_{b_2, e}$$



$$\Pr(L \in i, R \in j) = \sum_{b_1 \in \Omega_B} \sum_{b_2 \in \Omega_B} \sum_{l_1 \in \Omega_L} \sum_{l_2 \in \Omega_L} \sum_{e \in \Omega_E} P'(0, i)_{b_1, b_2} P'(i, i+1)_{b_2, l_1} P'(i+1, j)_{l_1, l_2} P'(j, j+1)_{l_2, e}$$



Algorithm

- Construct the state space
- Build the CTMC rate matrix
- Sum over all paths in the SCC graph

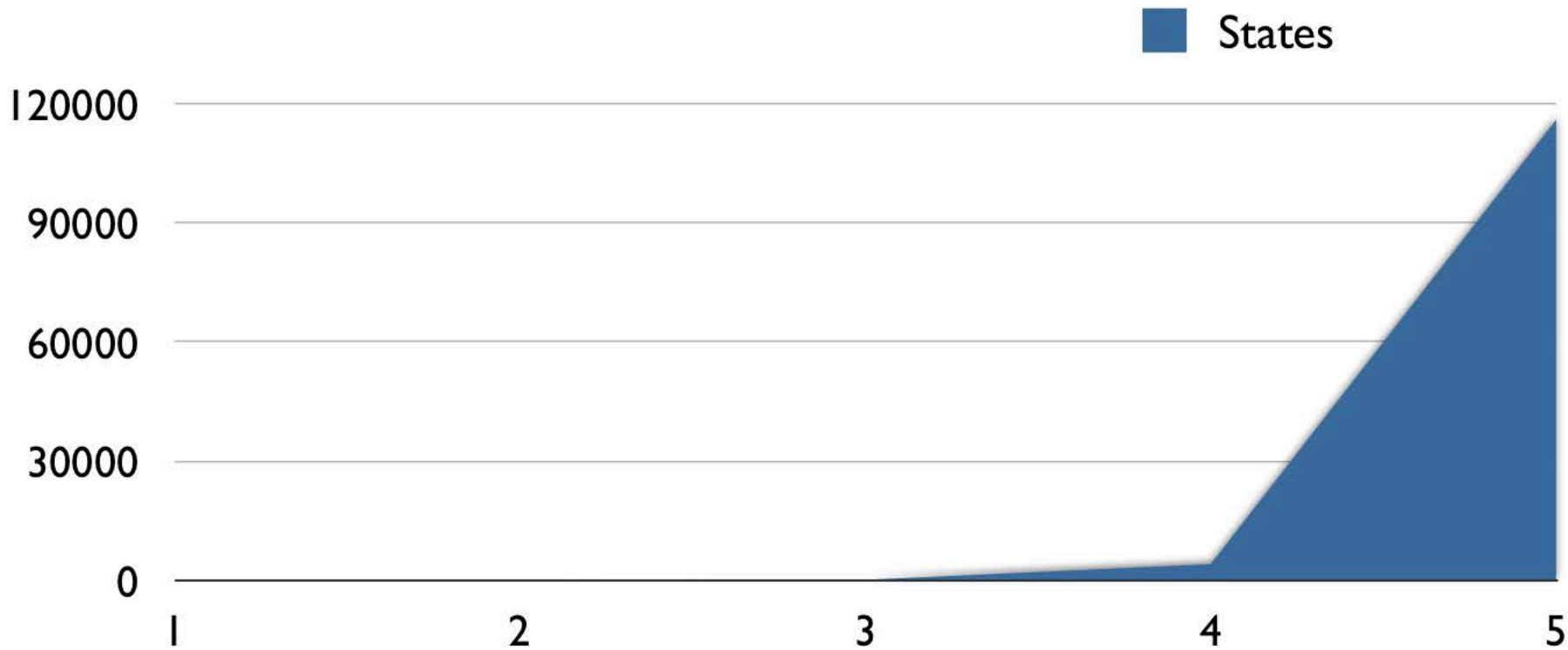
...there is a bit of a state
space explosion, though...

No genomes	No states	No transitions
1	2	2
2	15	44
3	203	1118
4	4140	35446
5	115975	1395582

$$Q = \begin{pmatrix} - & \dots & \dots & \dots & \dots & \dots \\ \dots & - & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & - \end{pmatrix}$$

$$(-c \ c \ 0 \ c \ 0 \ 0 \ 0 \ c \ 0 \ 0 \ 0)$$

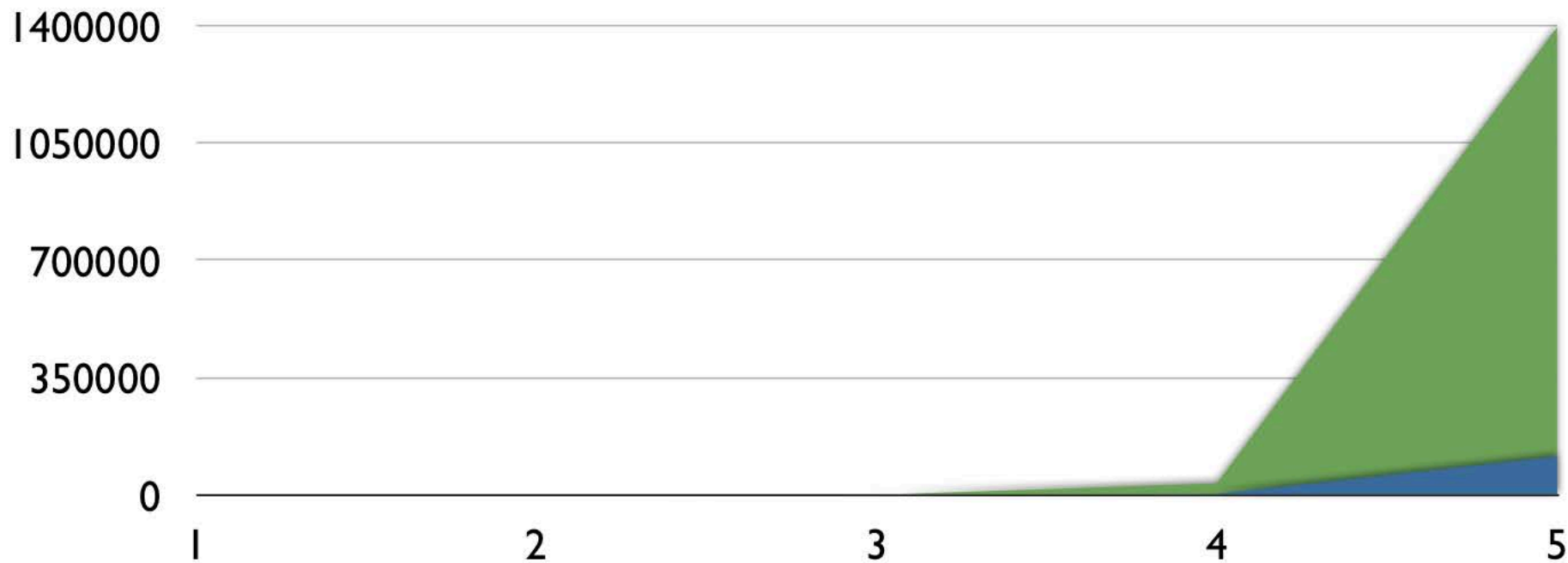
No genomes	No states	No transitions
1	2	2
2	15	44
3	203	1118
4	4140	35446
5	115975	1395582



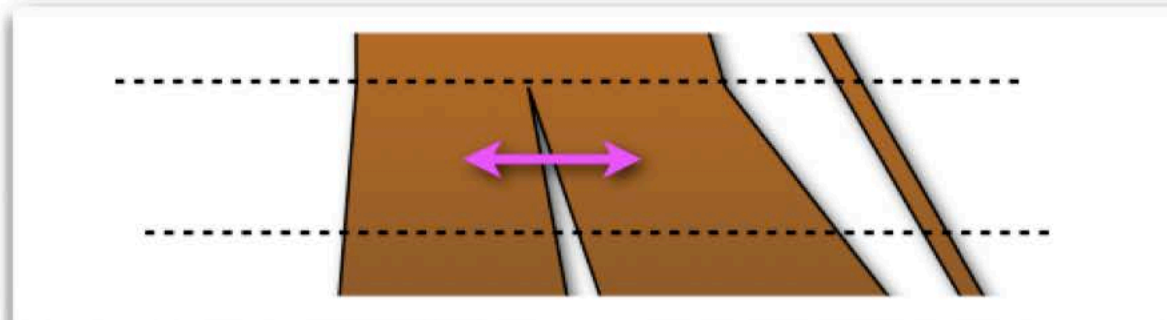
No genomes	No states	No transitions
1	2	2
2	15	44
3	203	1118
4	4140	35446
5	115975	1395582

■ States

■ Transitions

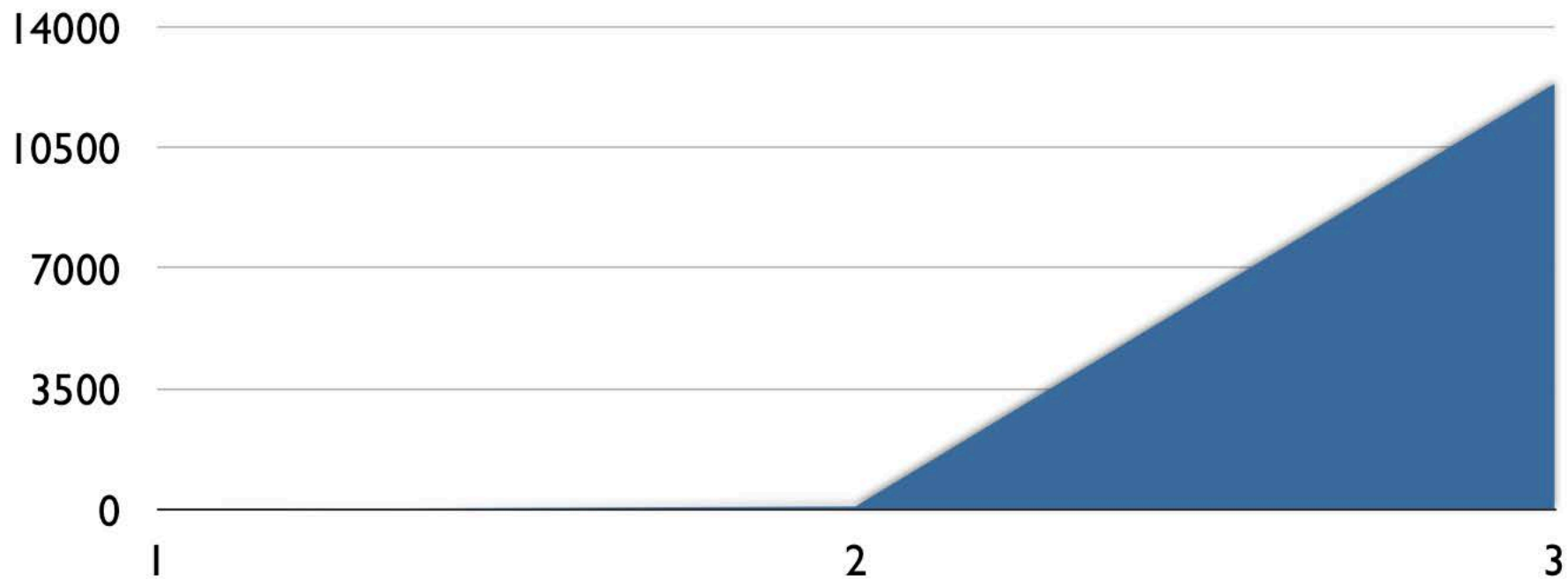


No genomes	No states	No transitions
1	2	2
2	94	466
3	12351	143886



No genomes	No states	No transitions
1	2	2
2	94	466
3	12351	143886

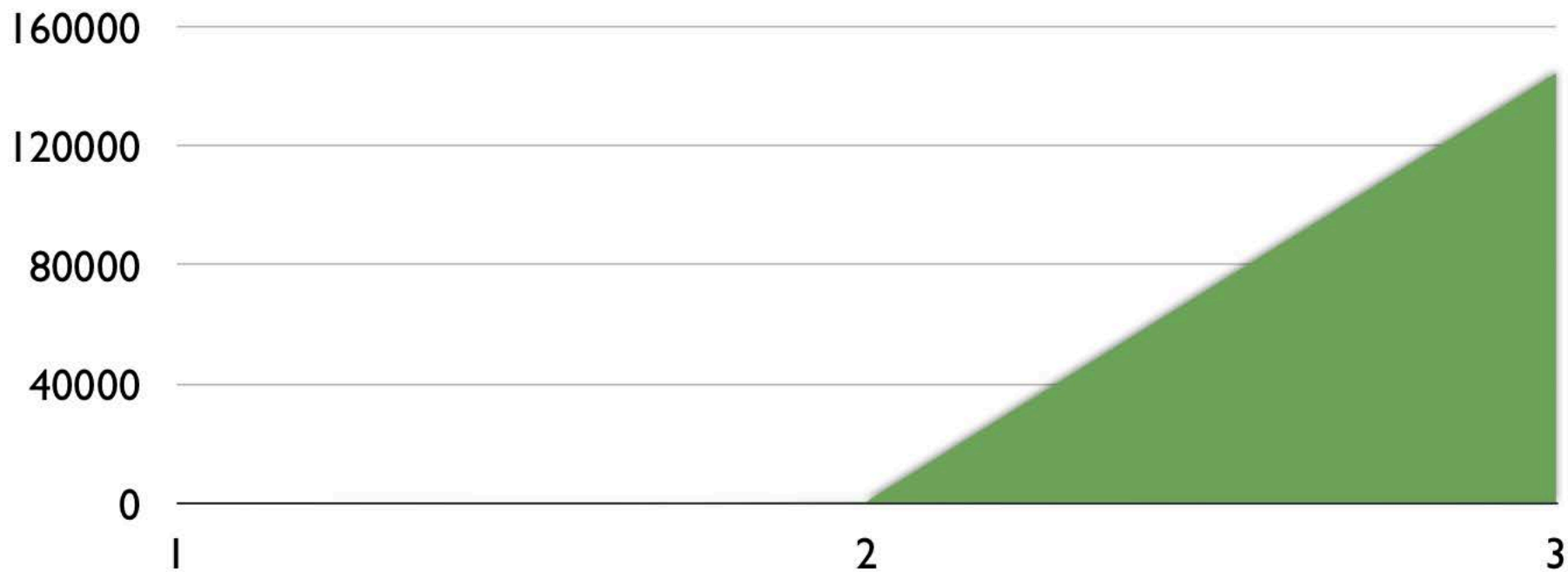
■ States

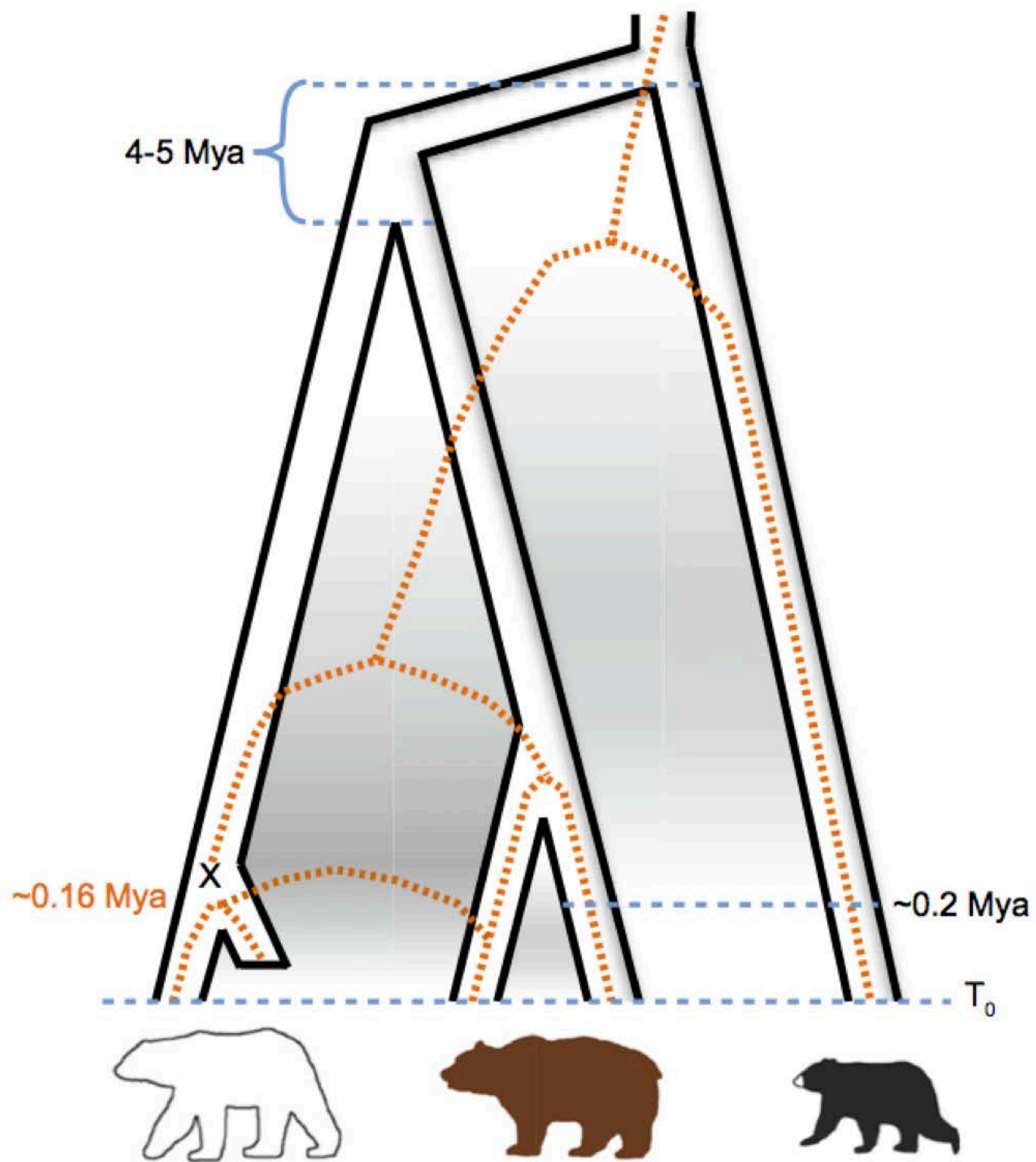


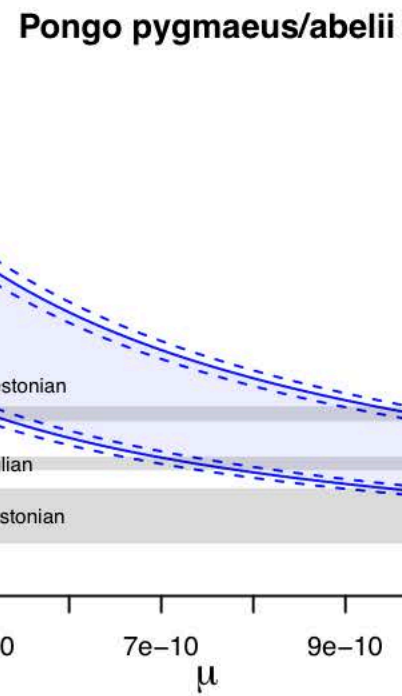
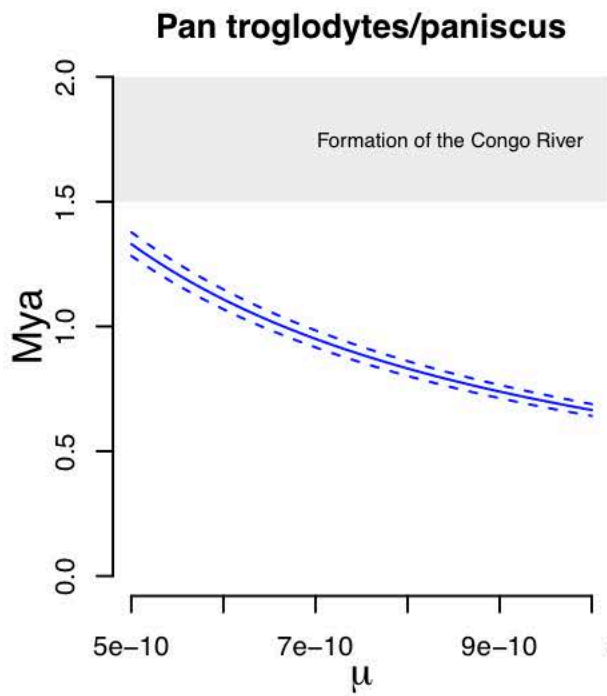
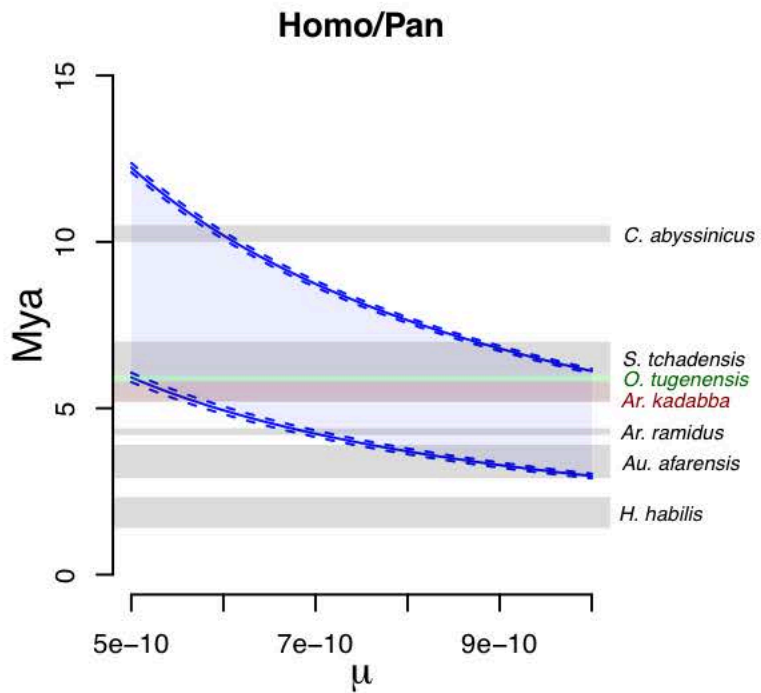
No genomes	No states	No transitions
1	2	2
2	94	466
3	12351	143886

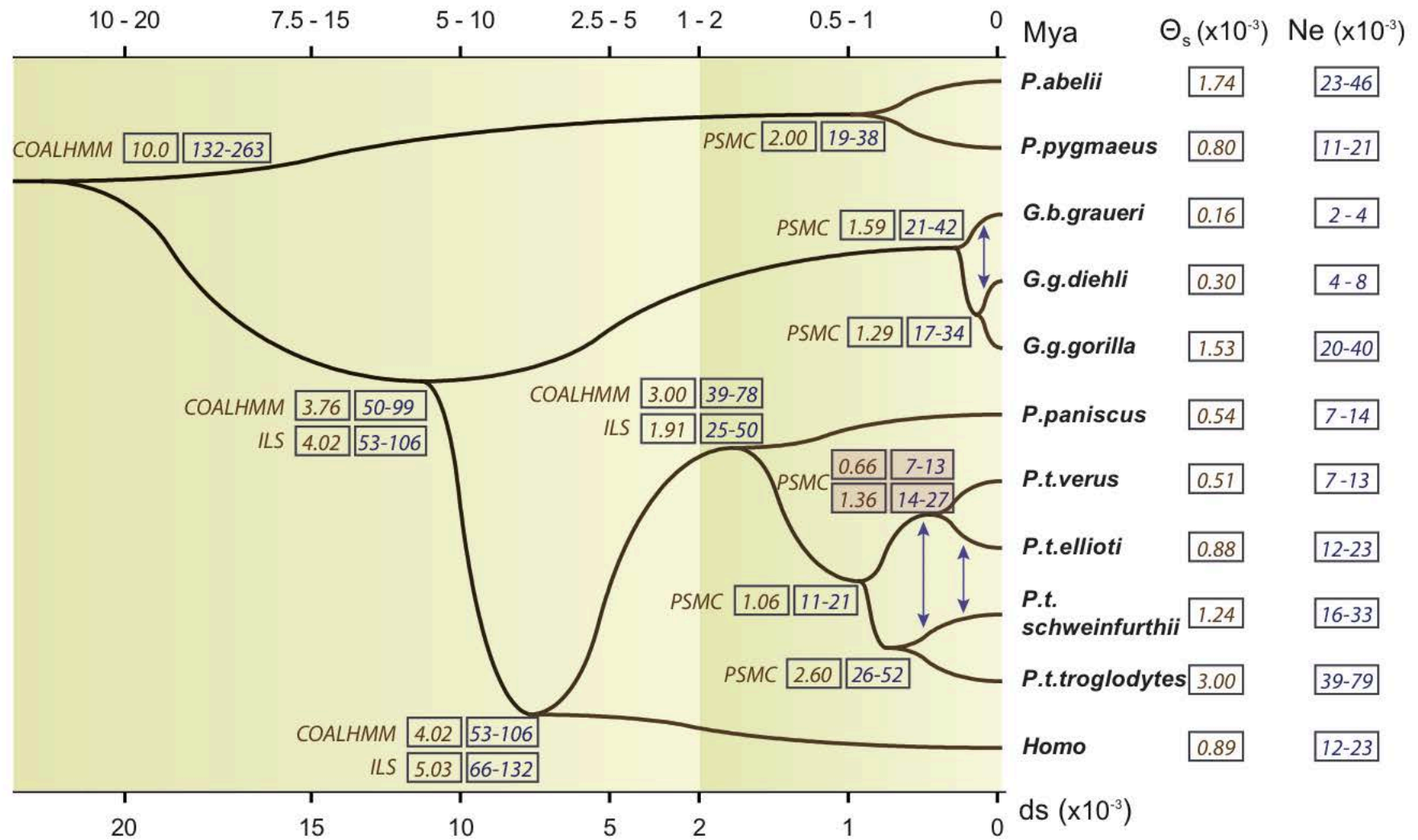
■ States

■ Transitions









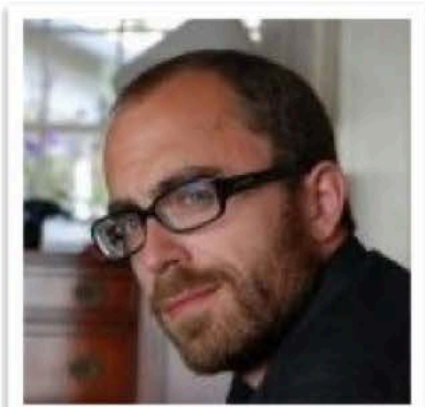


GENOME RESEARCH

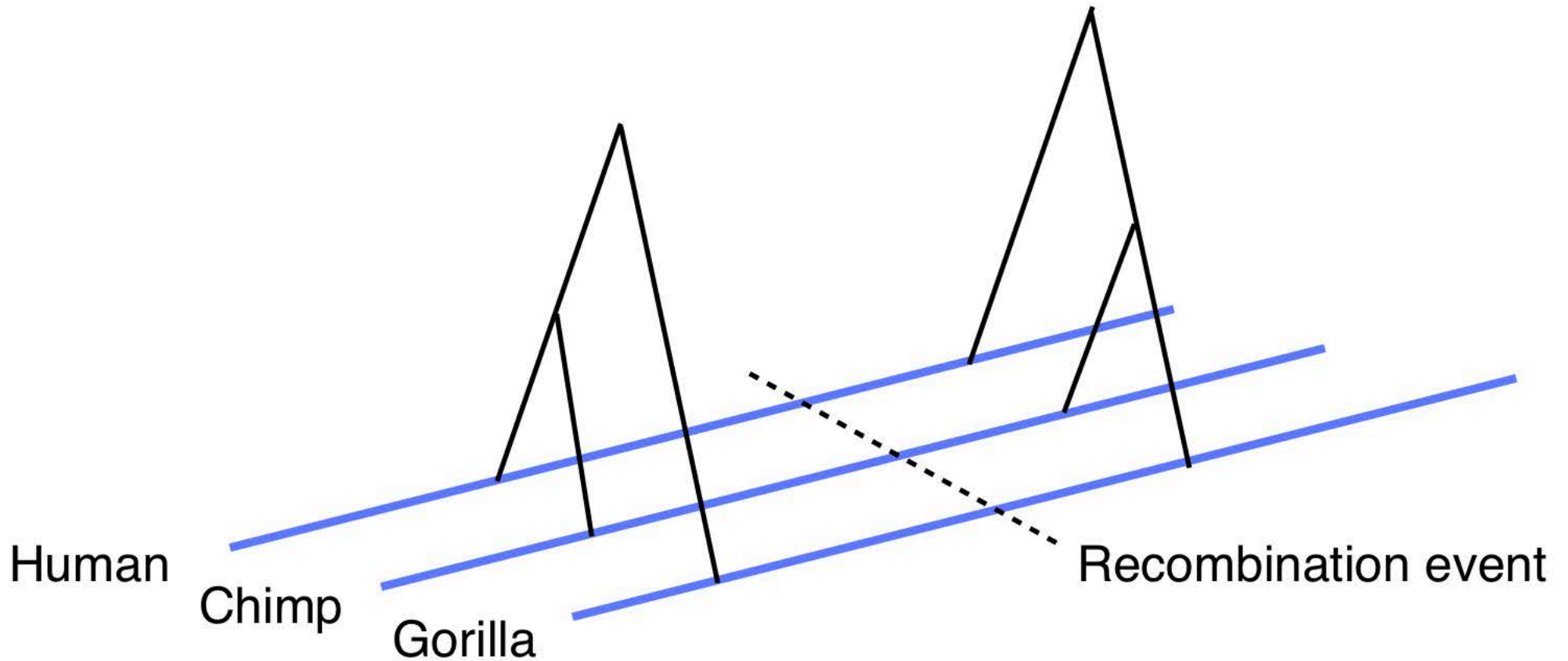
A fine-scale recombination map of the human-chimpanzee ancestor reveals faster change in humans than in chimpanzees and a strong impact of GC-biased gene conversion

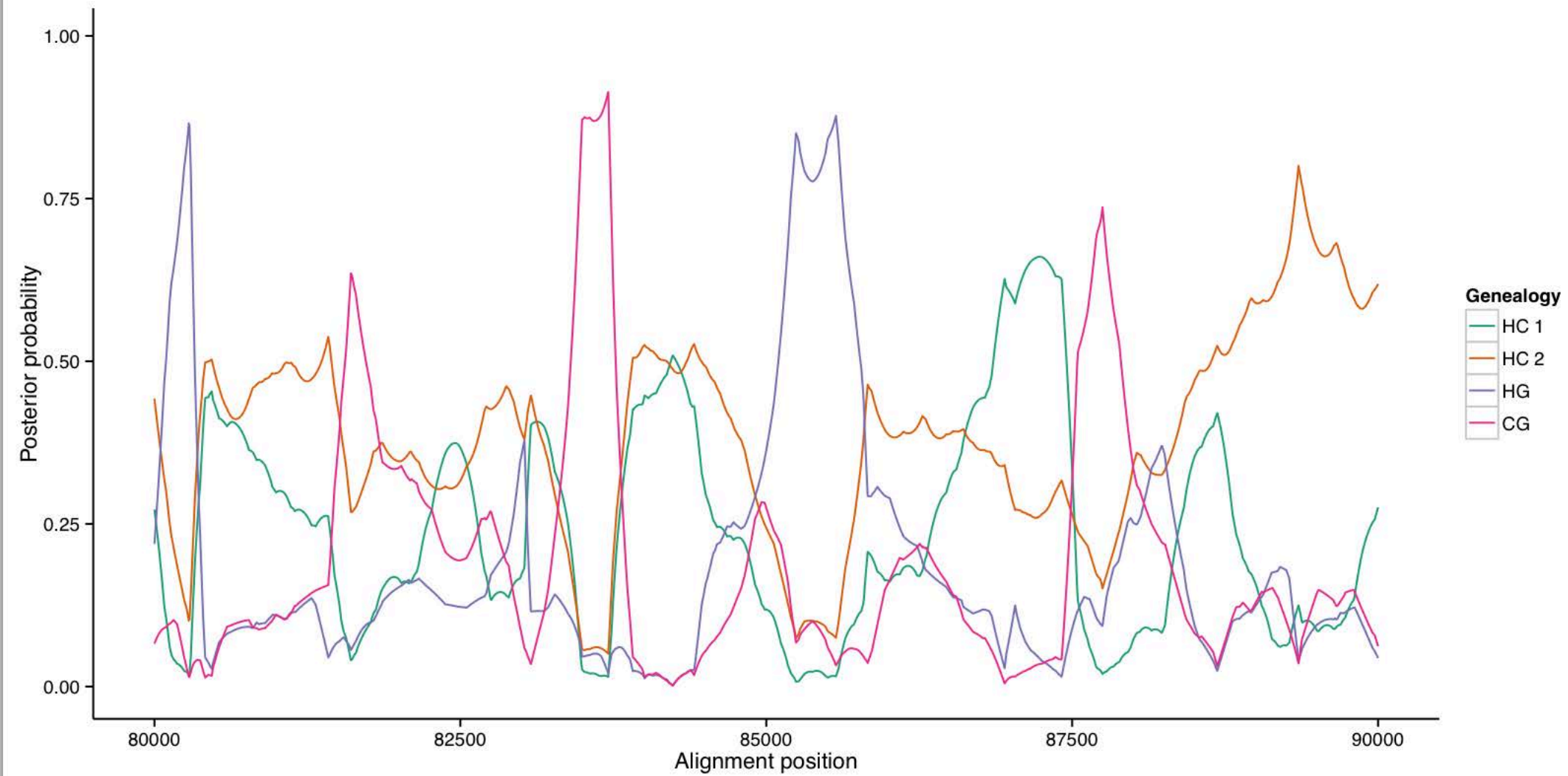
Kasper Munch, Thomas Mailund, Julien Y Dutheil, et al. **Mikkel H. Schierup**

Genome Res. published online November 4, 2013
Access the most recent version at doi:[10.1101/gr.158469.113](https://doi.org/10.1101/gr.158469.113)



Inferring recombination events

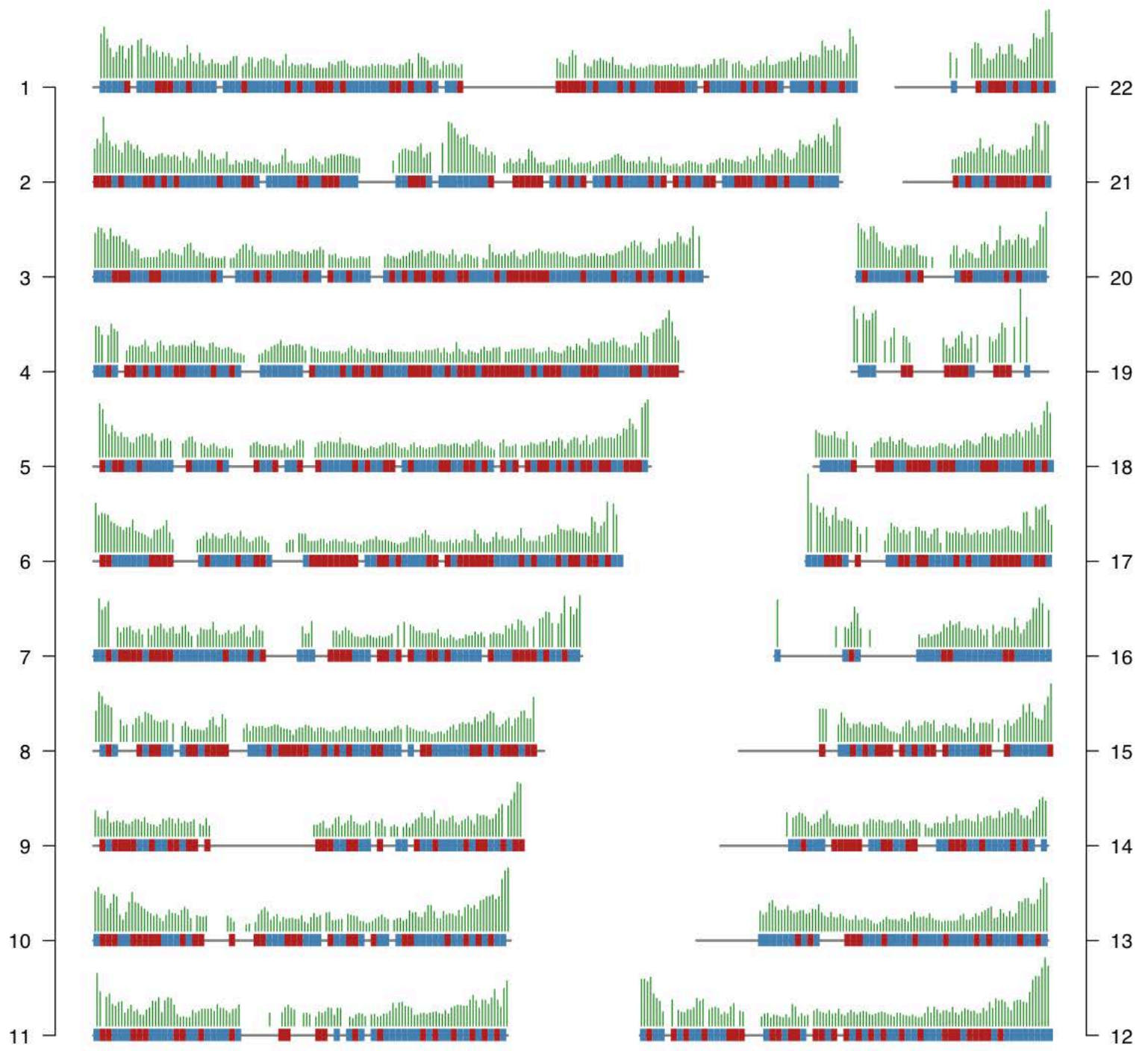




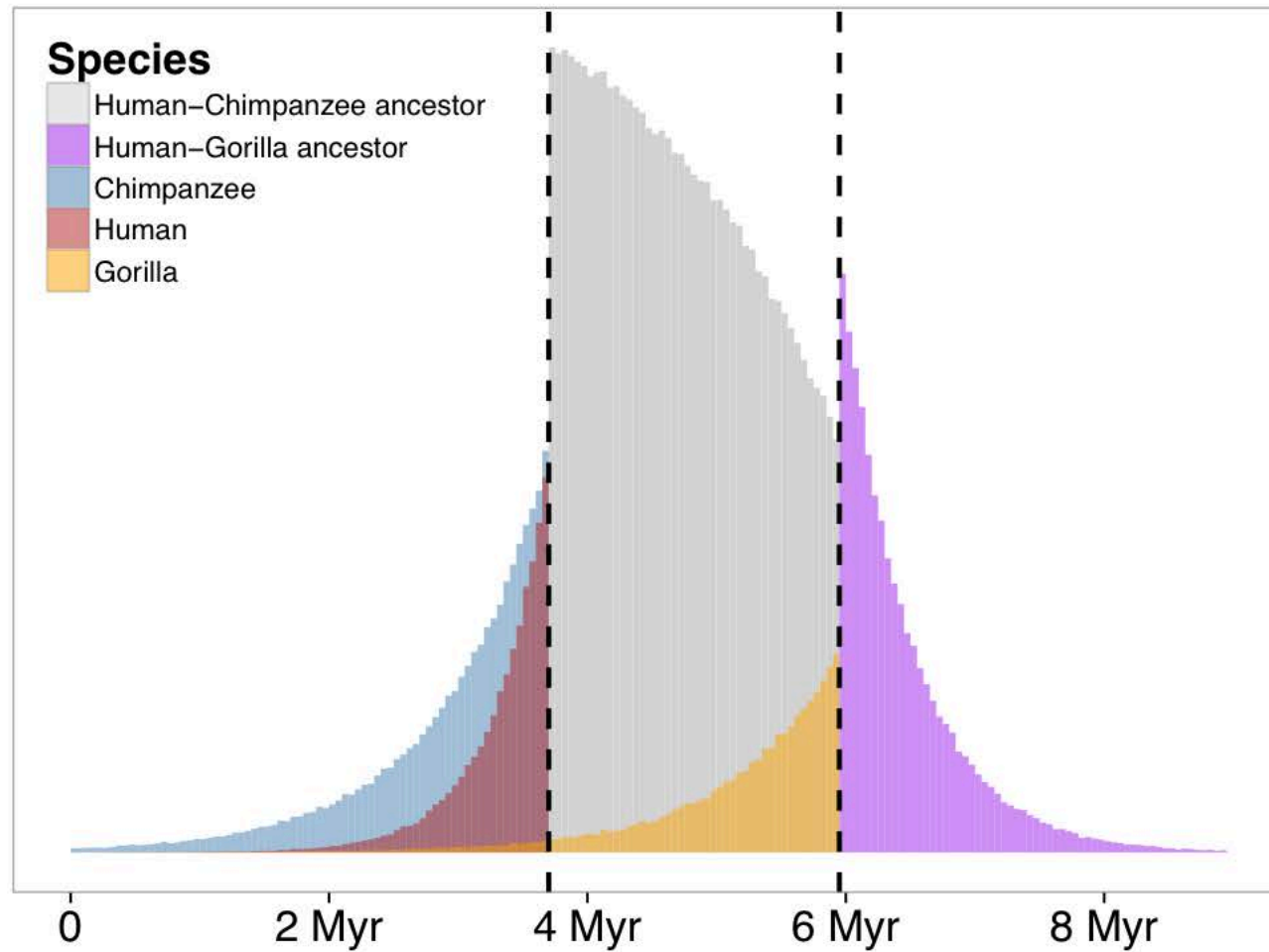
5 cM/Mb



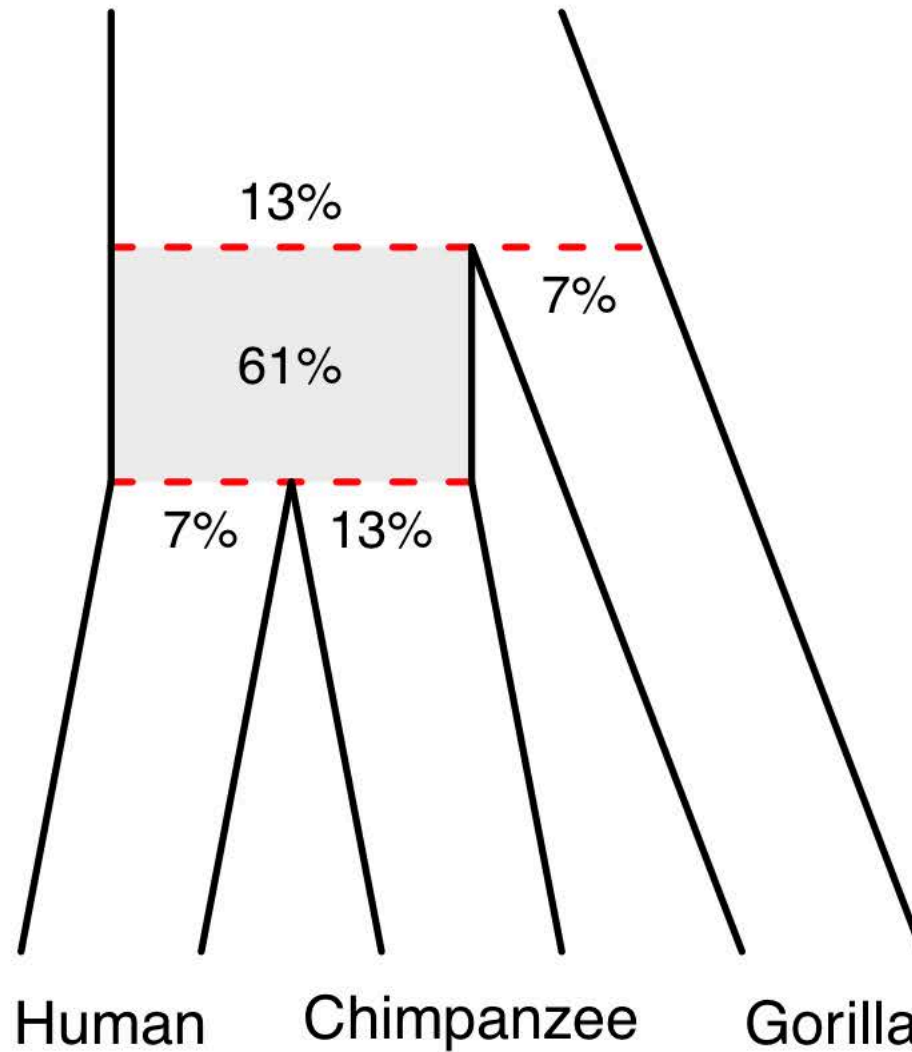
Chromosomes



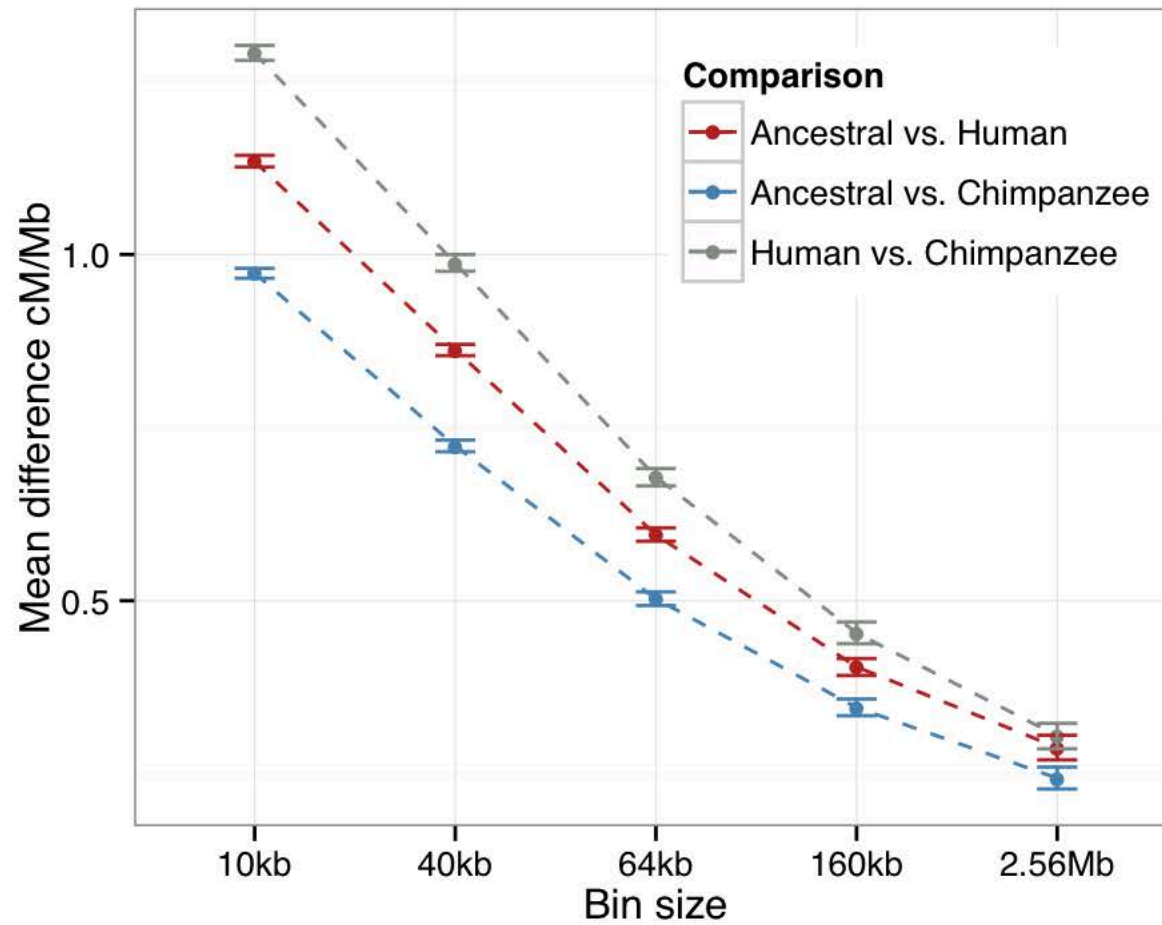
Distribution of called recombination events



Relative contributions



Larger difference to humans





"That's all Folks!"