

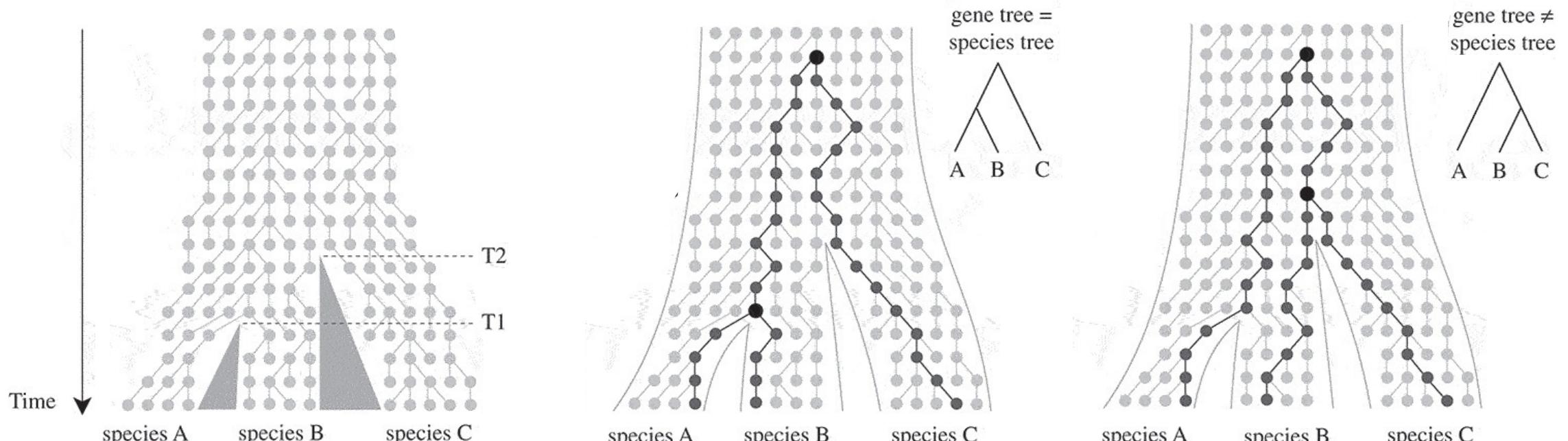
ILSMC: a phylogenetically aware SMC to infer the evolutionary history of species

Iker Rivas-González

Overview

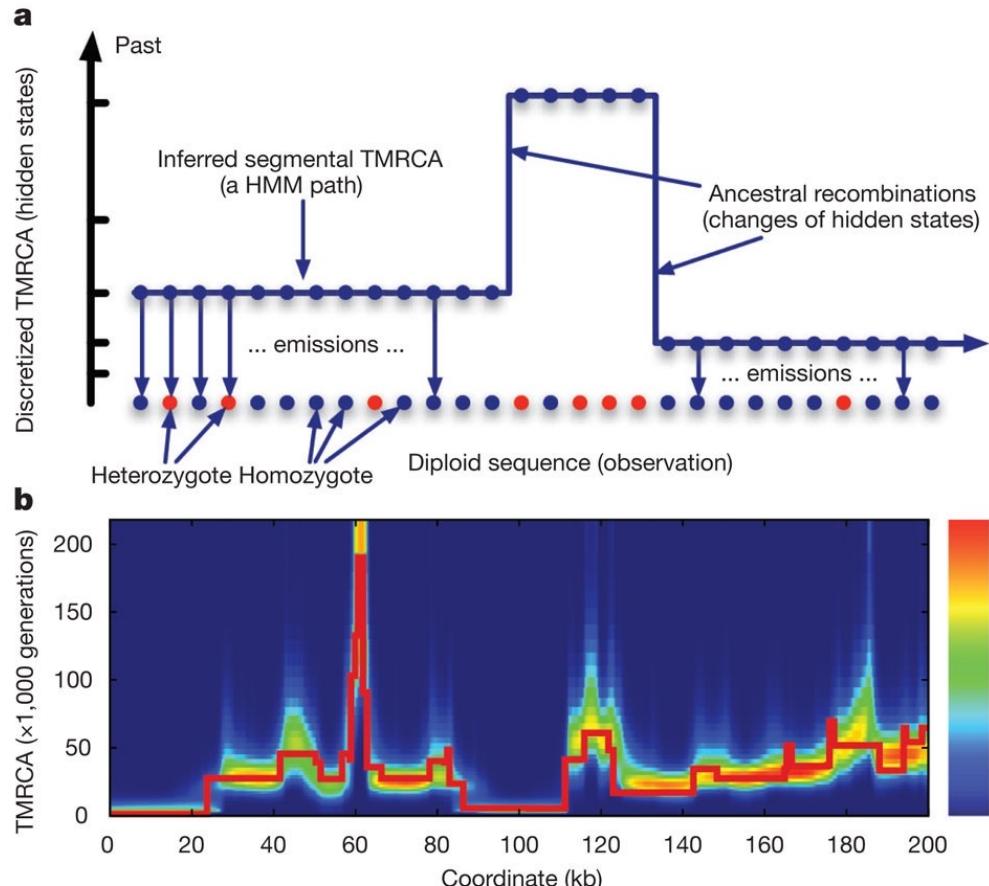
- Incomplete lineage sorting (ILS)
- HMMs: CoalHMM and SMCs
- ILSMC
 - The hidden states
 - The CTMCs
 - The transition probability matrix
 - Simulating from the model
- Discussion

Incomplete lineage sorting (ILS)

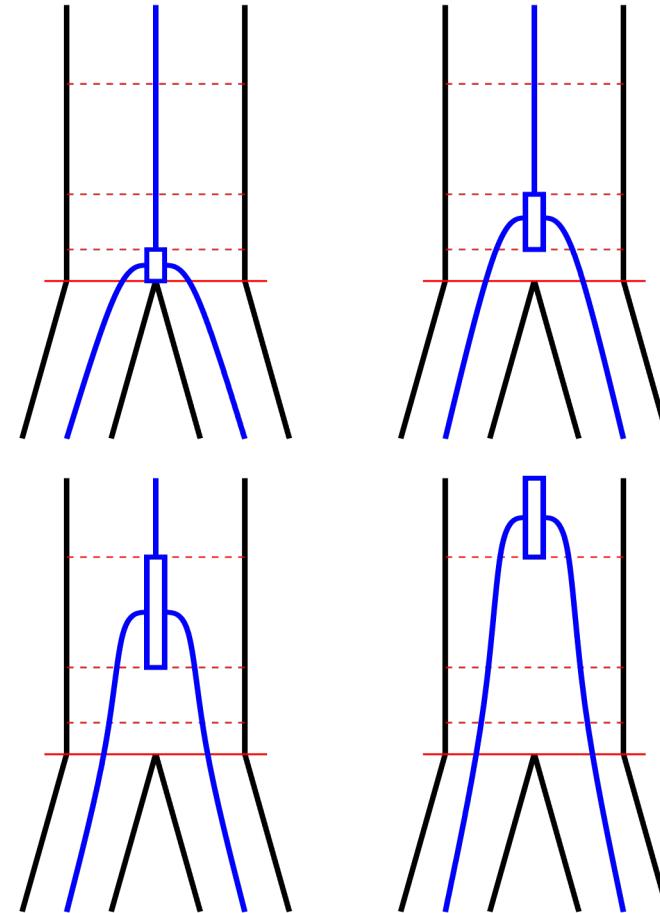


Modified from Leliaert et al. (2013)

HMMs



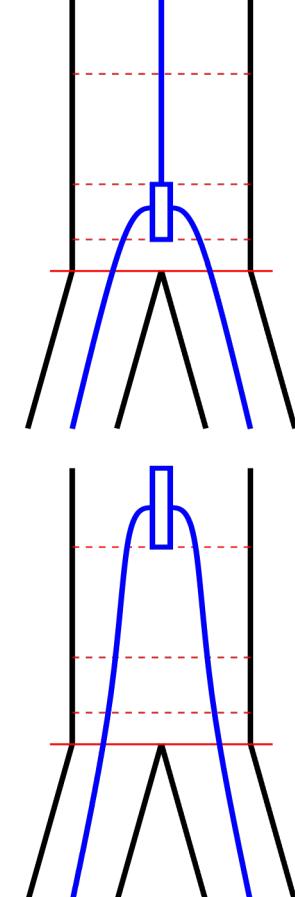
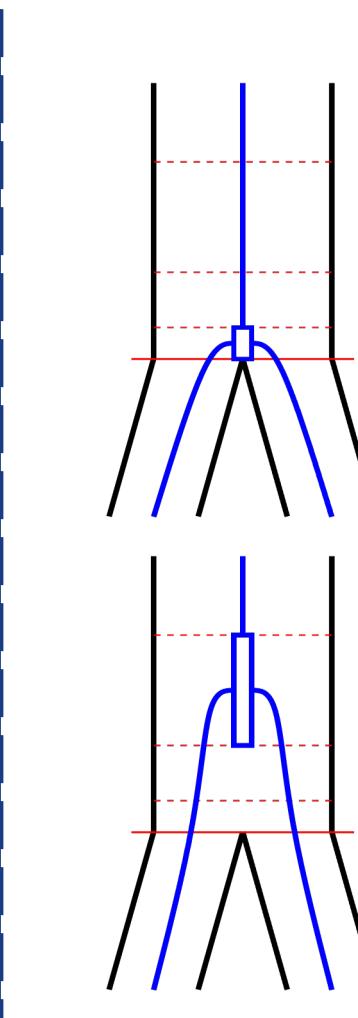
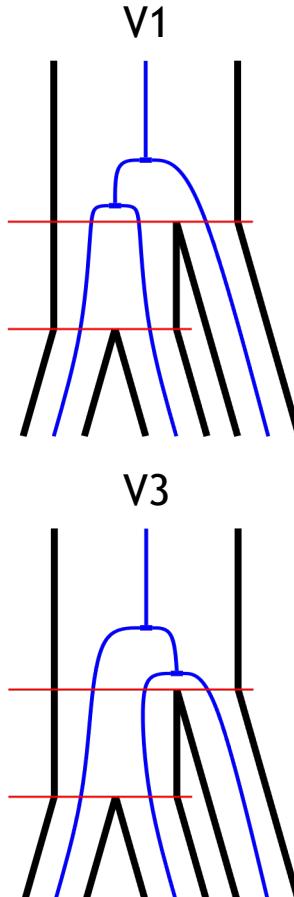
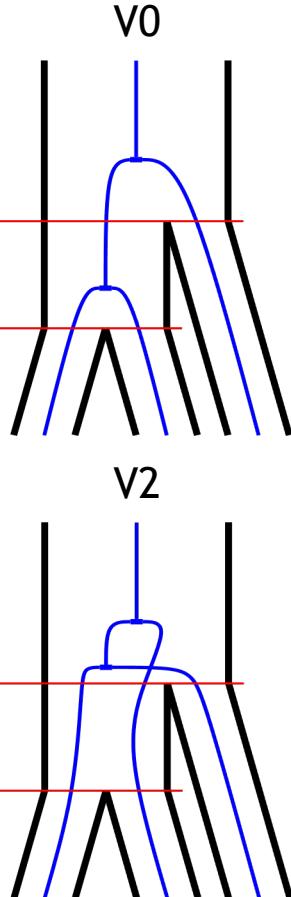
Li and Durbin (2011)



SMCs

HMMs

CoalHMM

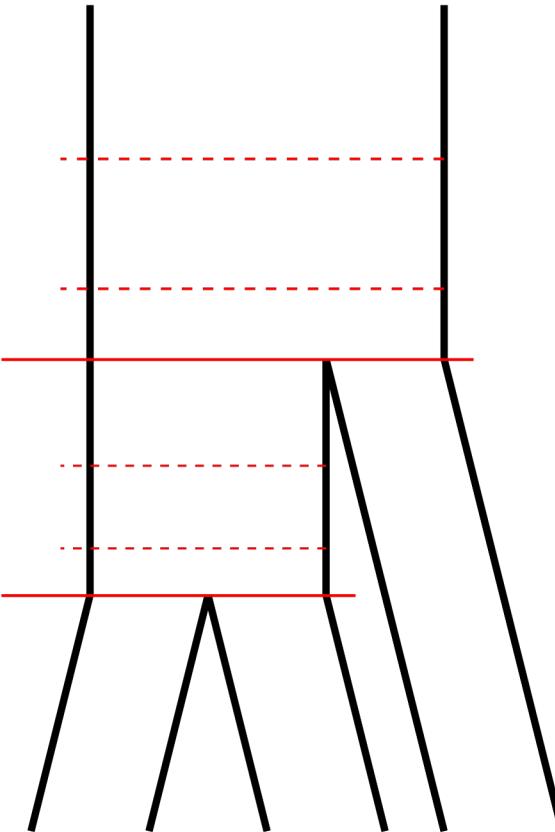


SMCs

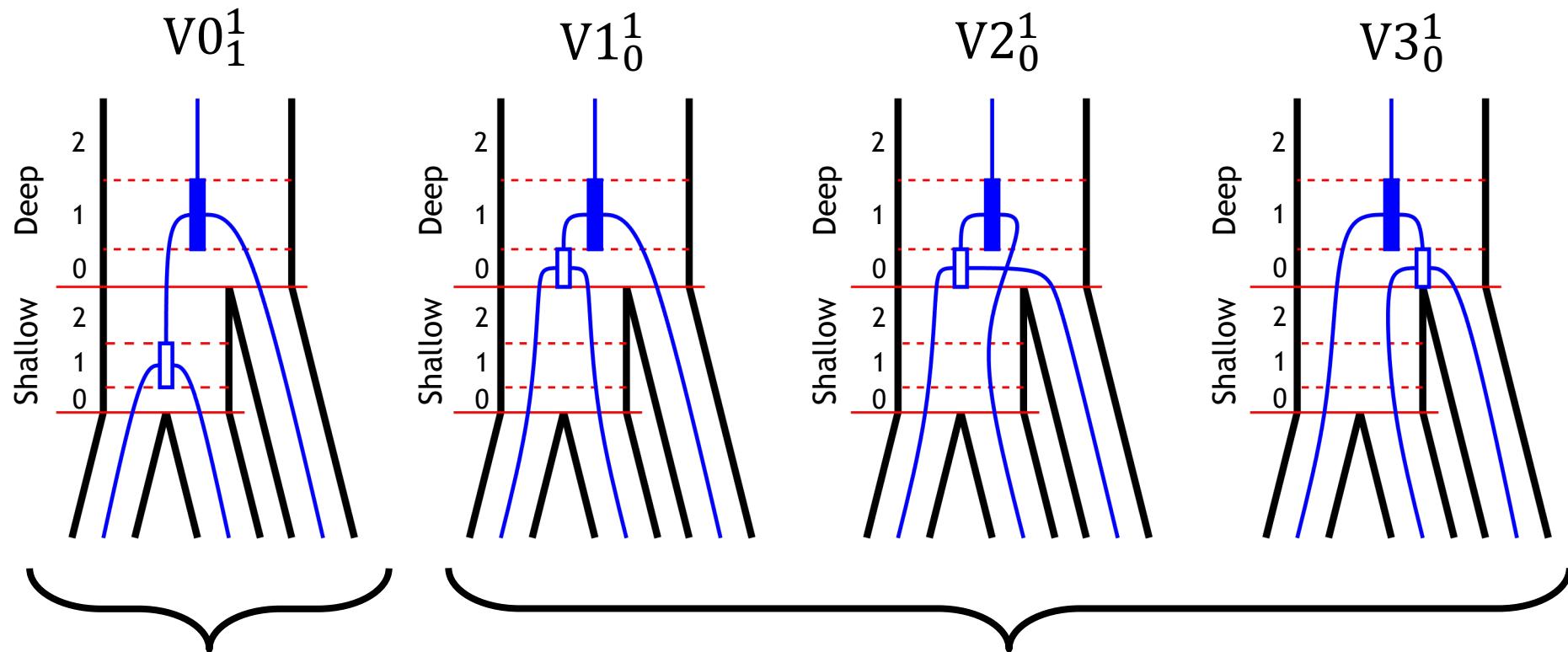
ILSMC

Extract the ILS
information
CoalHMM-style

Parameterize with
discretized time
intervals SMC-style



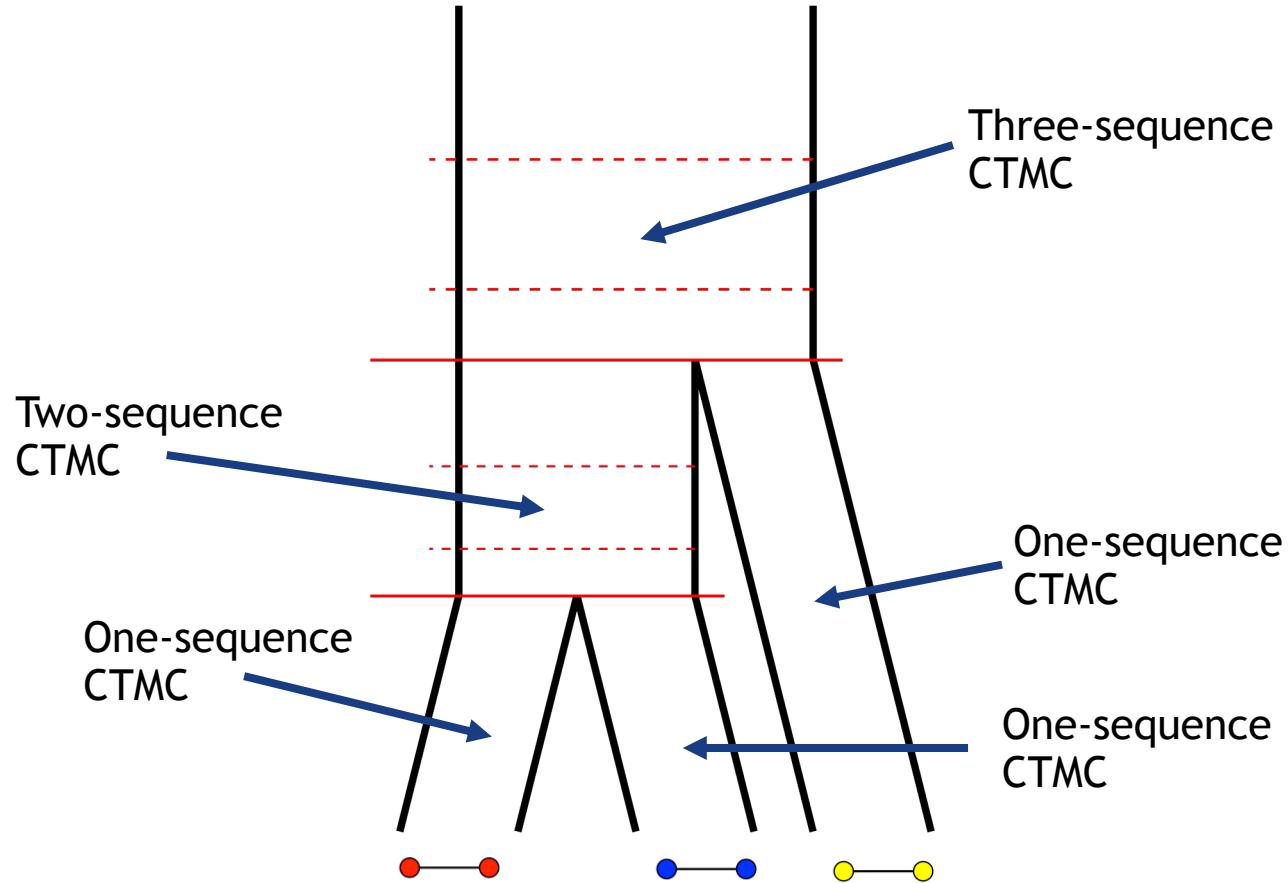
The hidden states



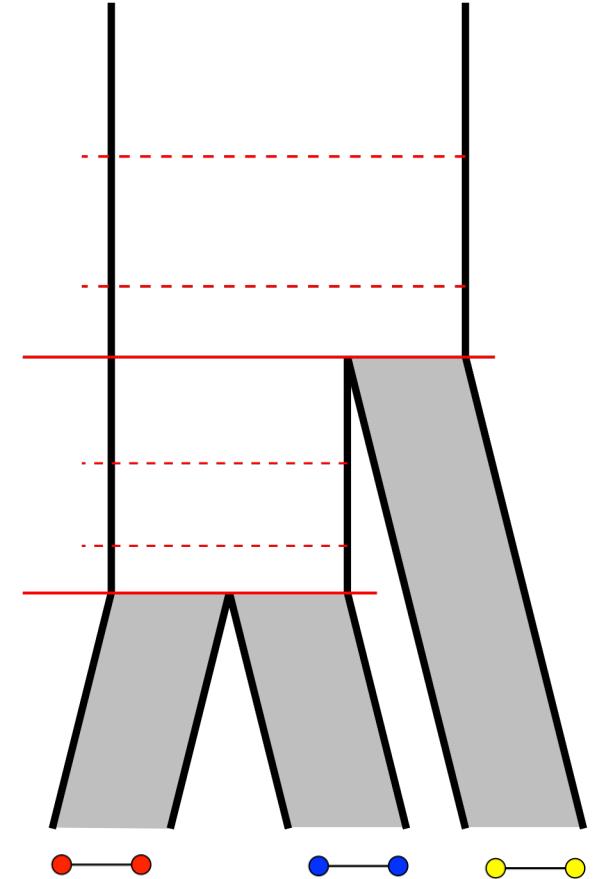
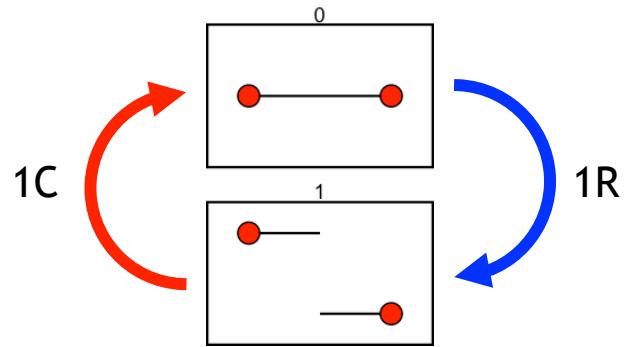
$V0^1_1$
interval of 2nd coalescent (deep)
interval of 1st coalescent (shallow)

VX^1_0
interval of 2nd coalescent (deep)
interval of 1st coalescent (deep)

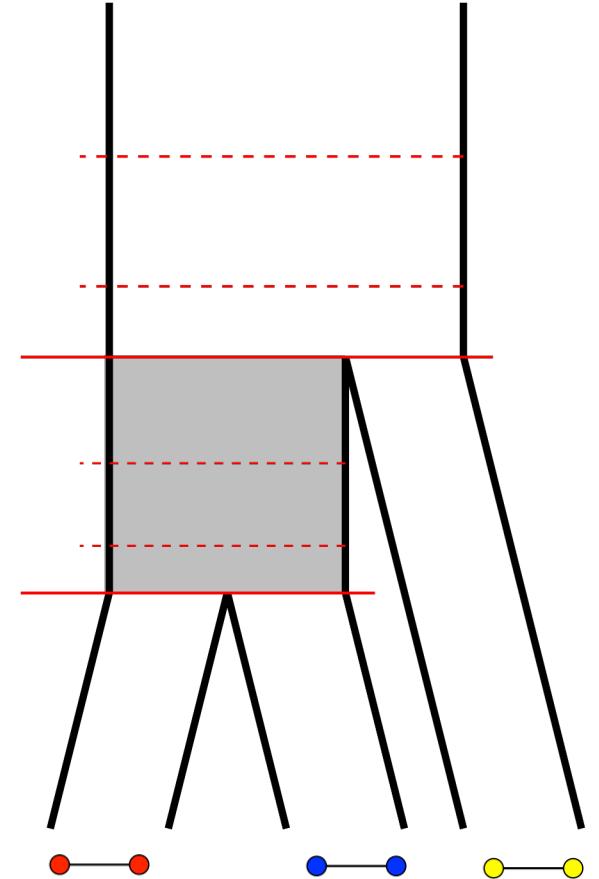
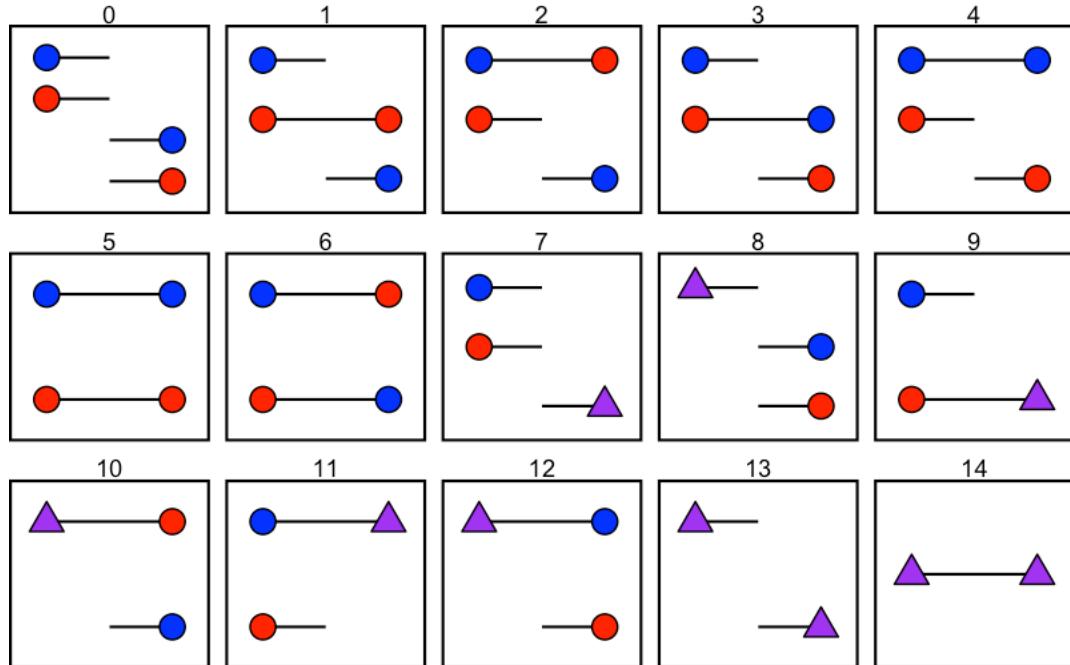
The CTMCs



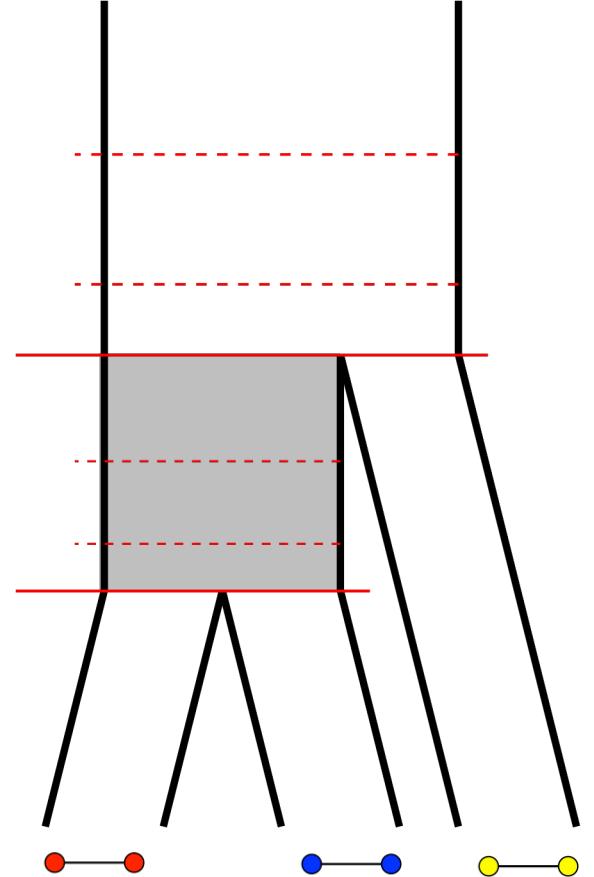
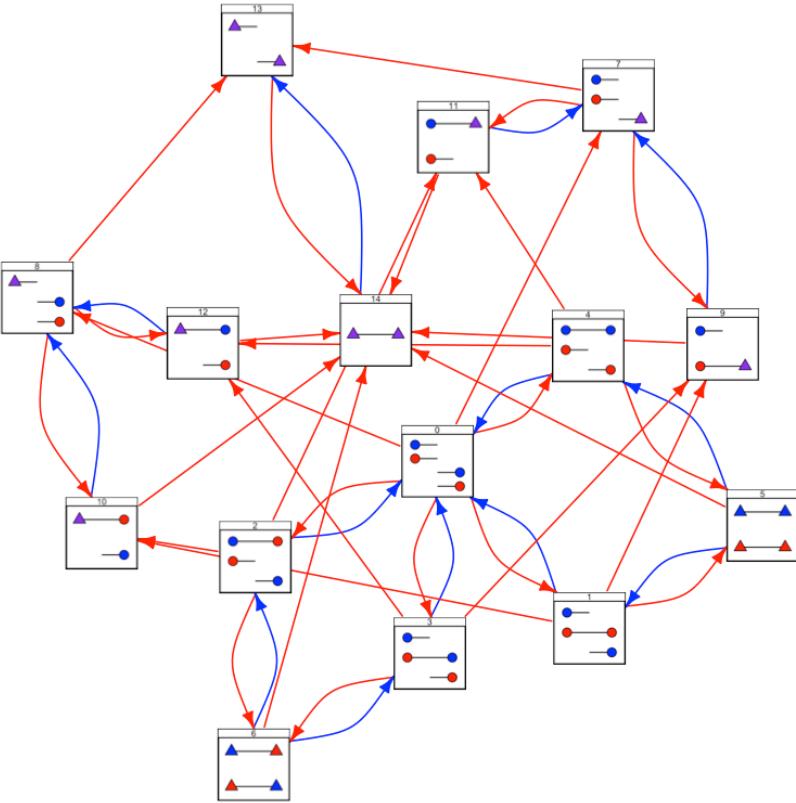
The CTMCs



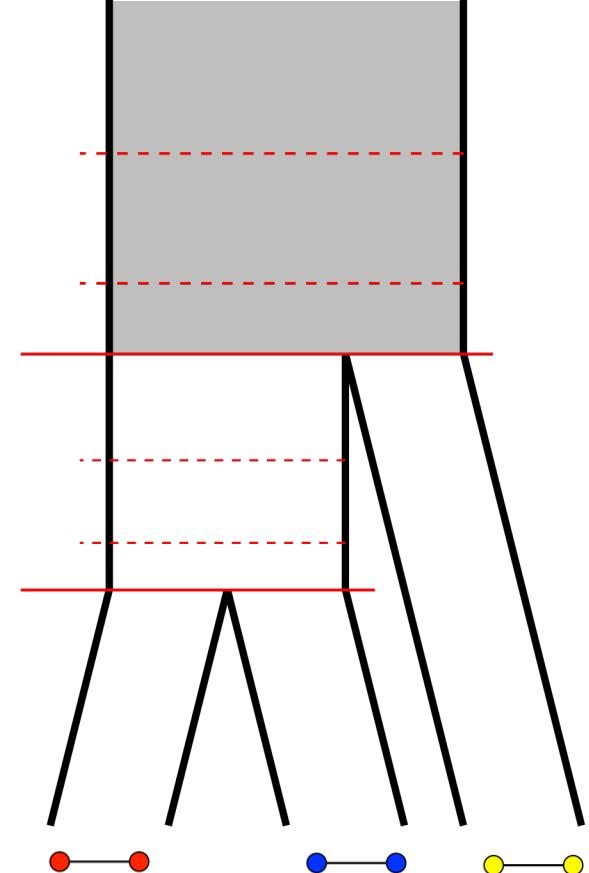
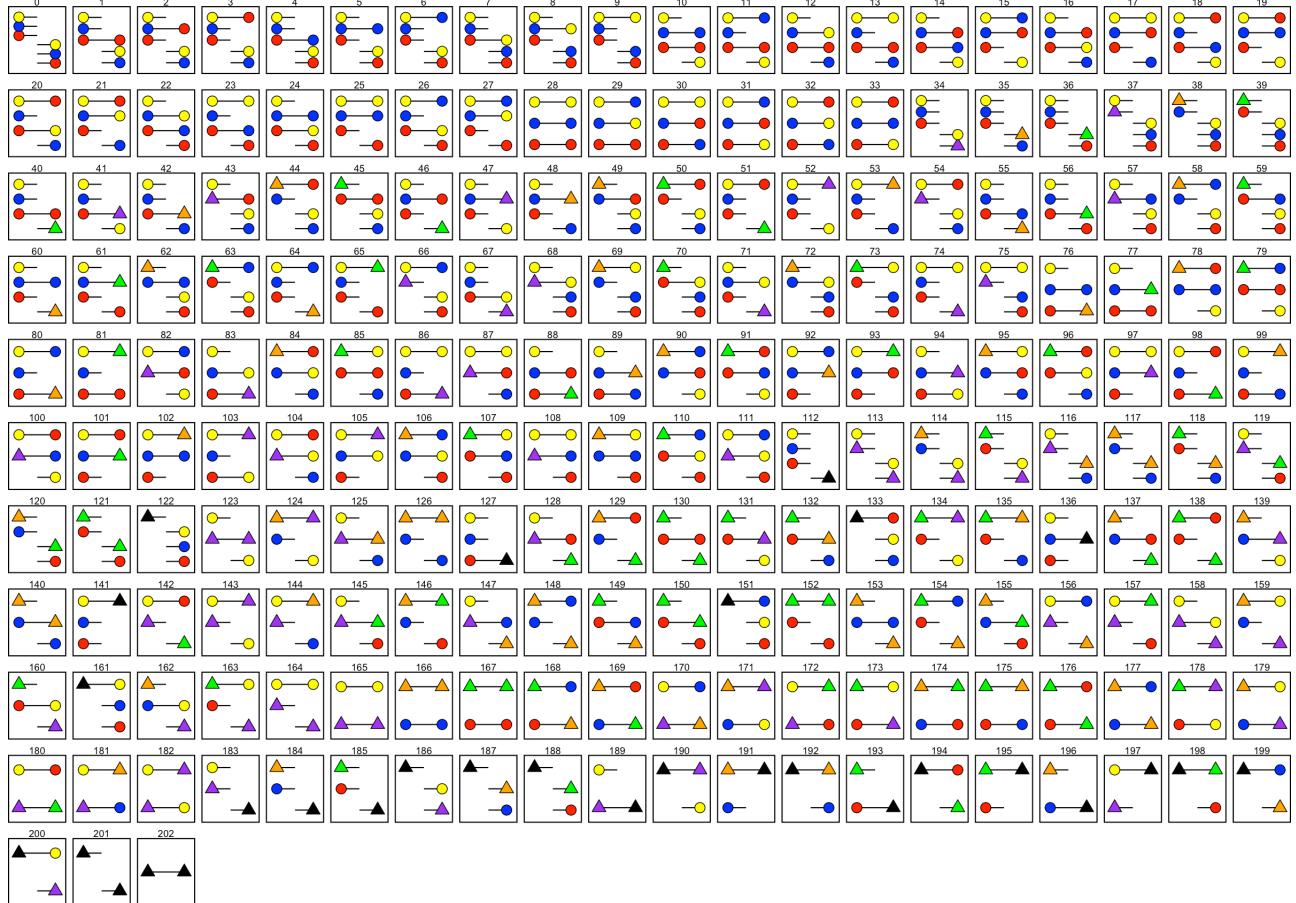
The CTMCs



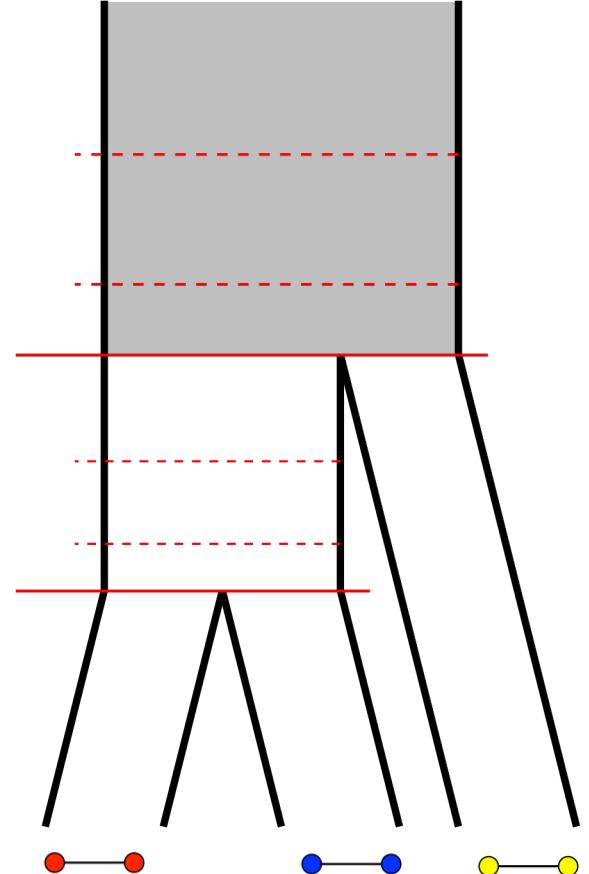
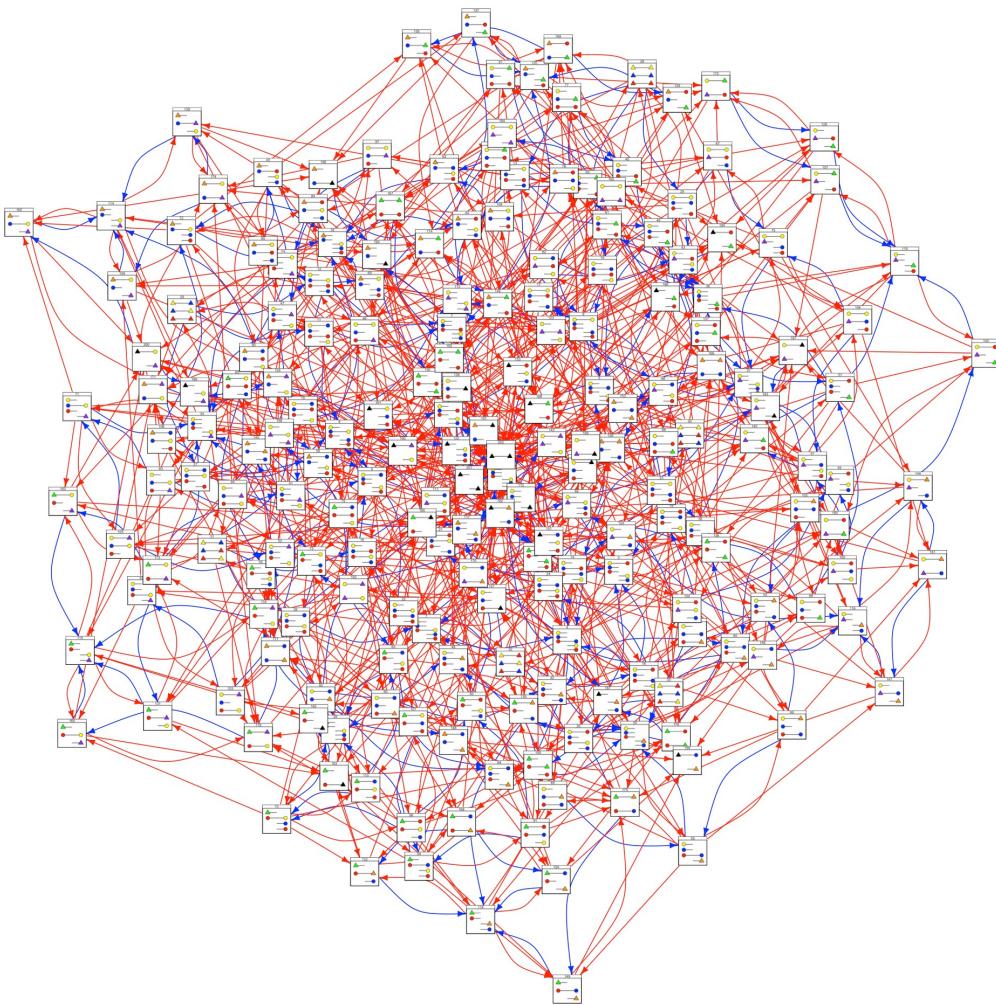
The CTMCs



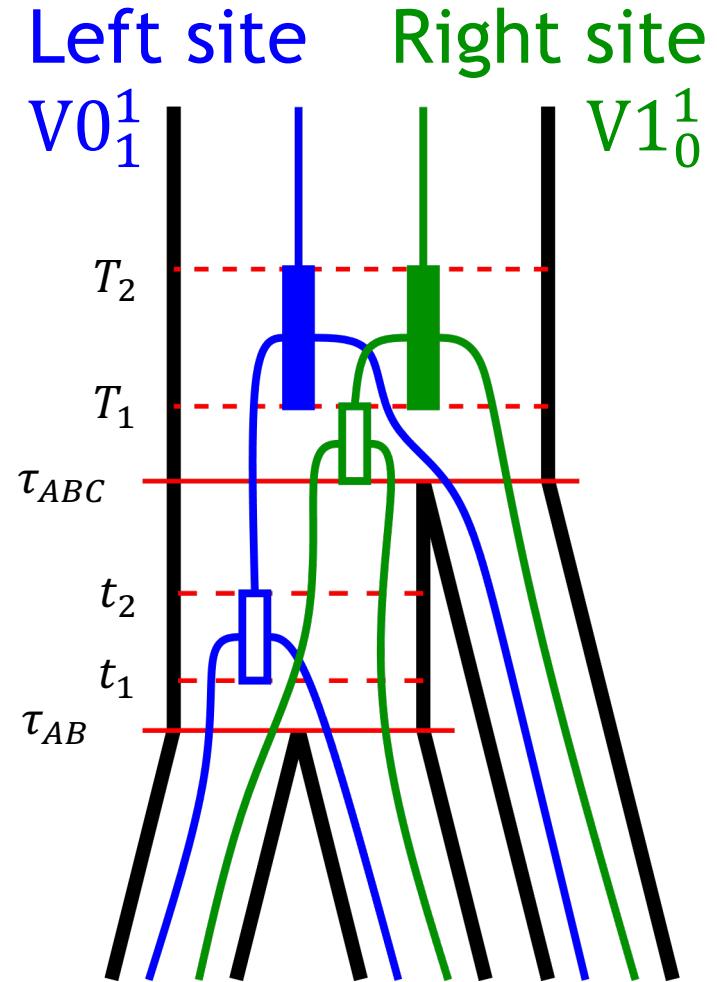
The CTMCs



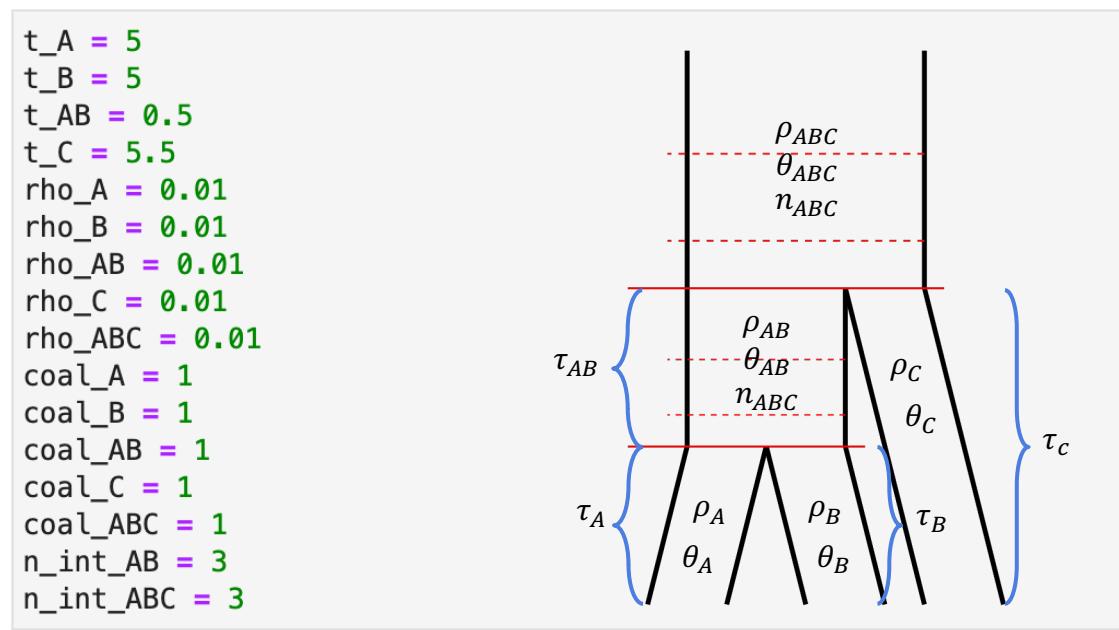
The CTMCs



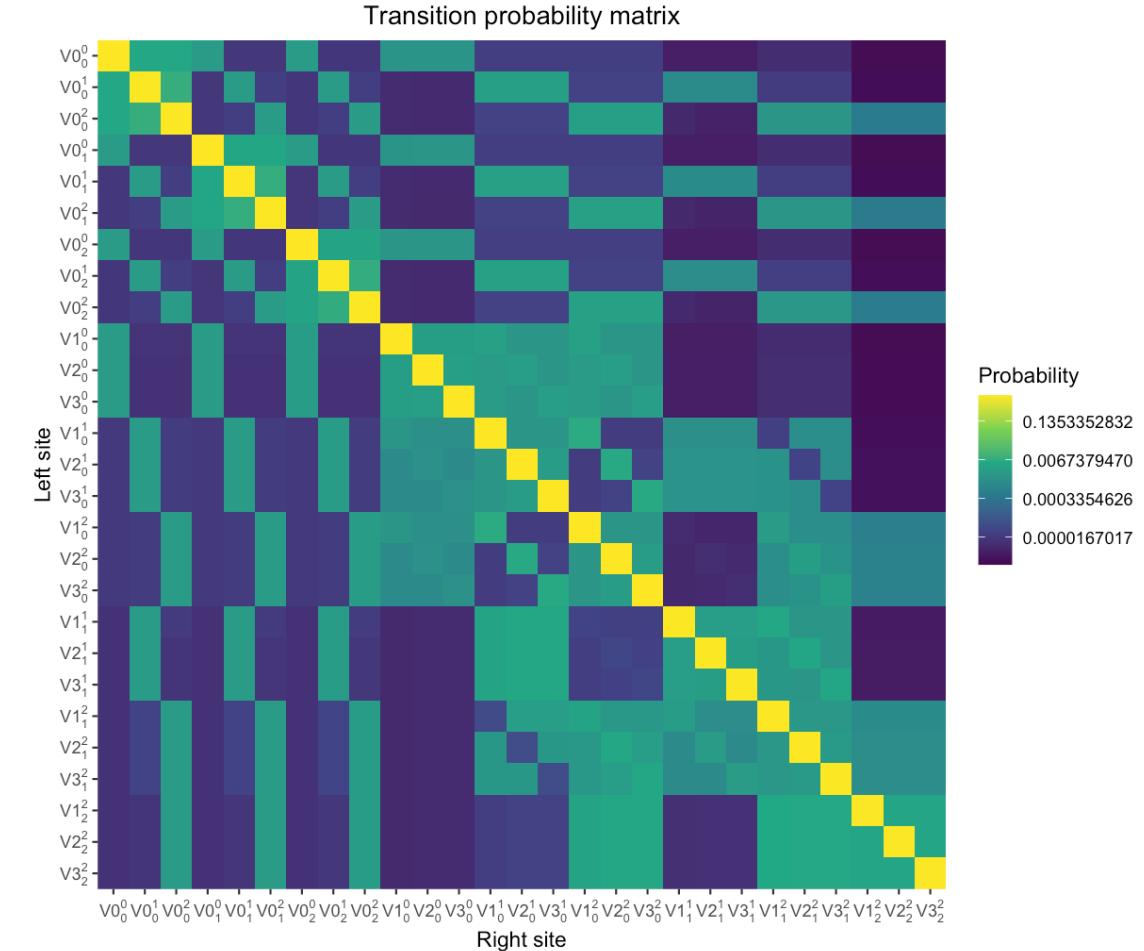
The transition probability matrix



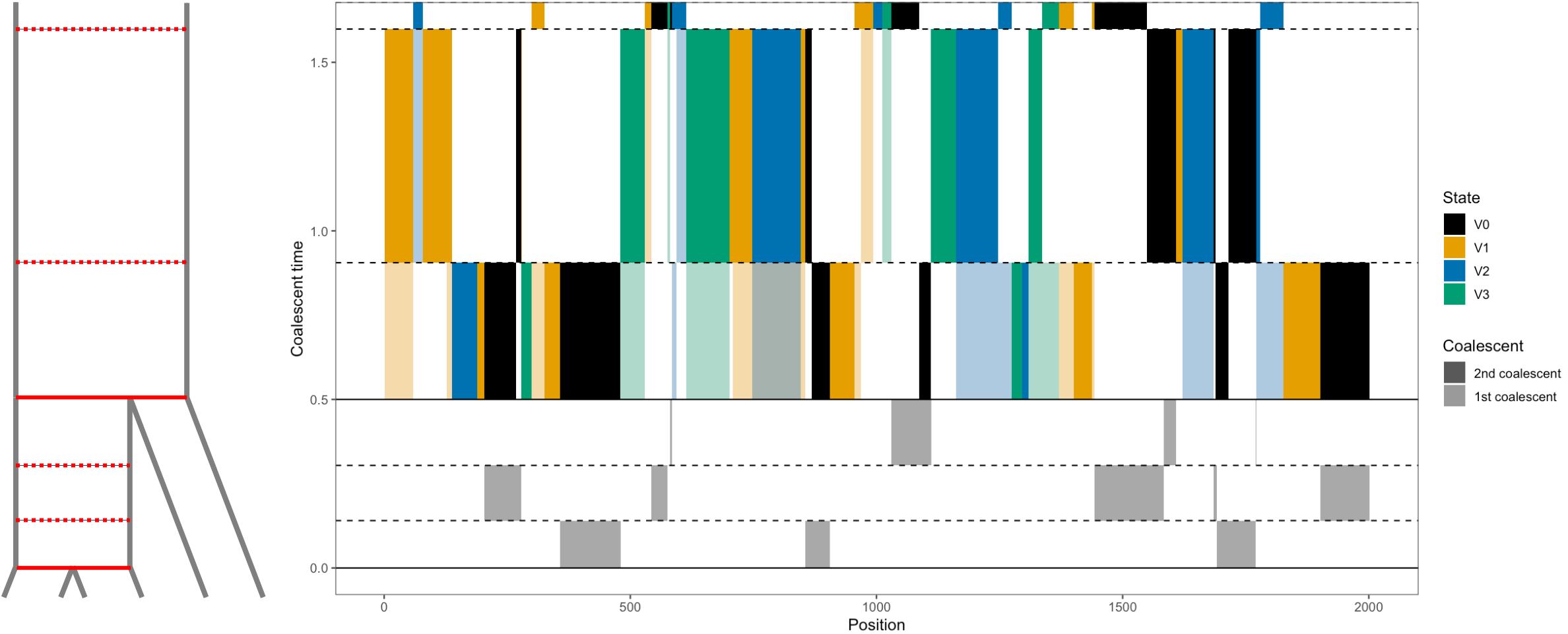
The transition probability matrix



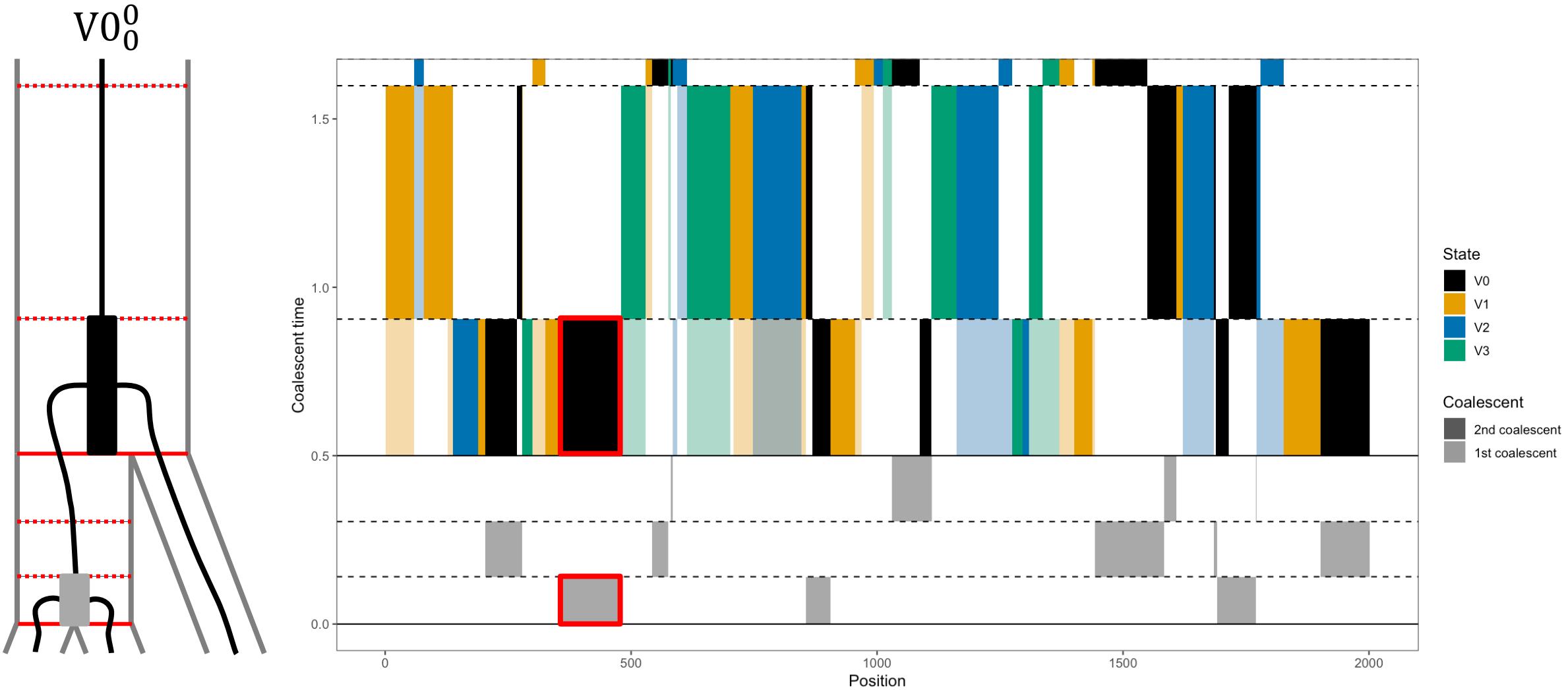
```
array_np = get_HMM_trans_mat(
    t_A,      t_B,      t_AB,      t_C,
    rho_A,    rho_B,    rho_AB,    rho_C,    rho_ABC,
    coal_A,   coal_B,   coal_AB,   coal_C,   coal_ABC,
    n_int_AB, n_int_ABC
)
```



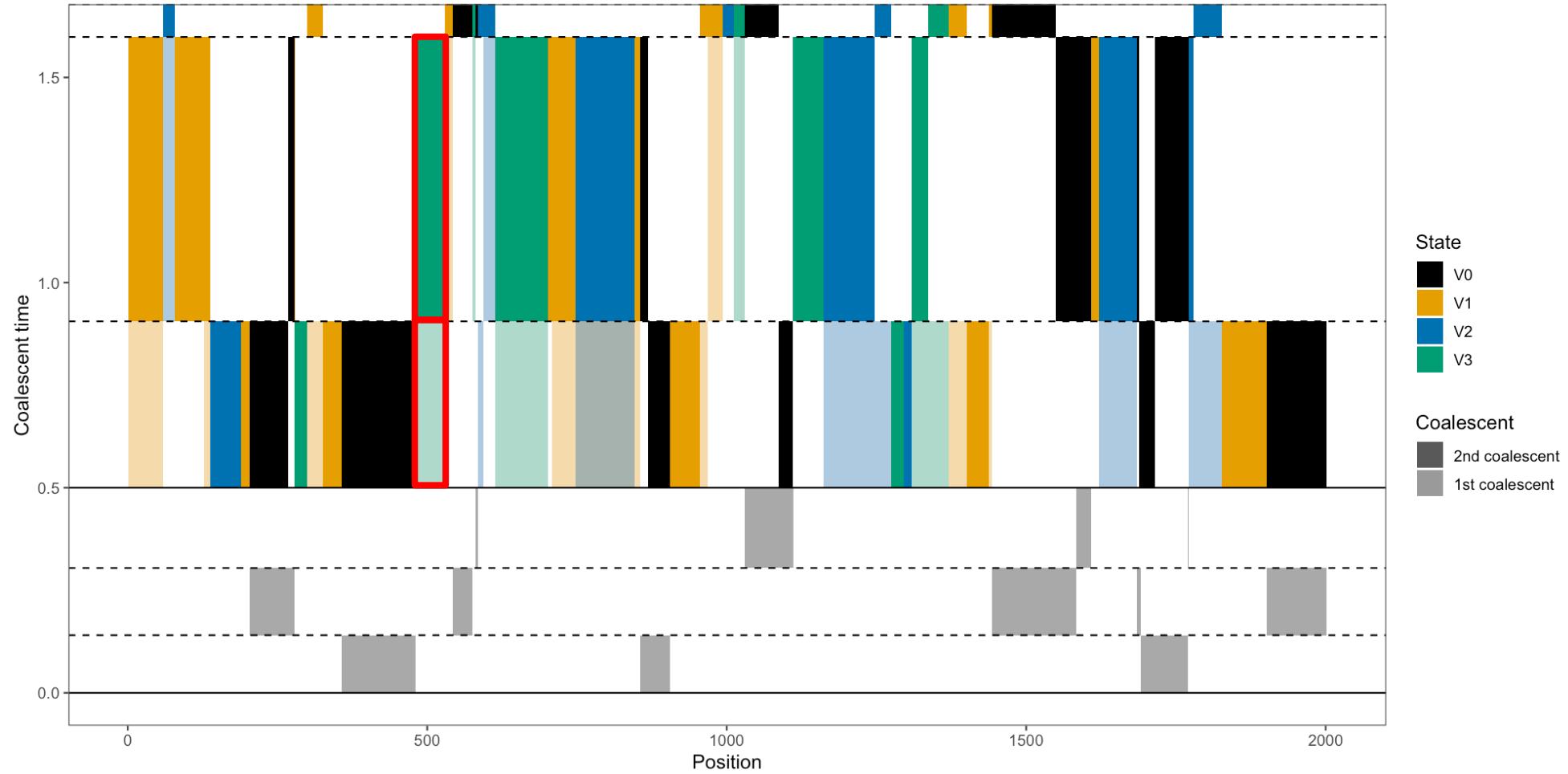
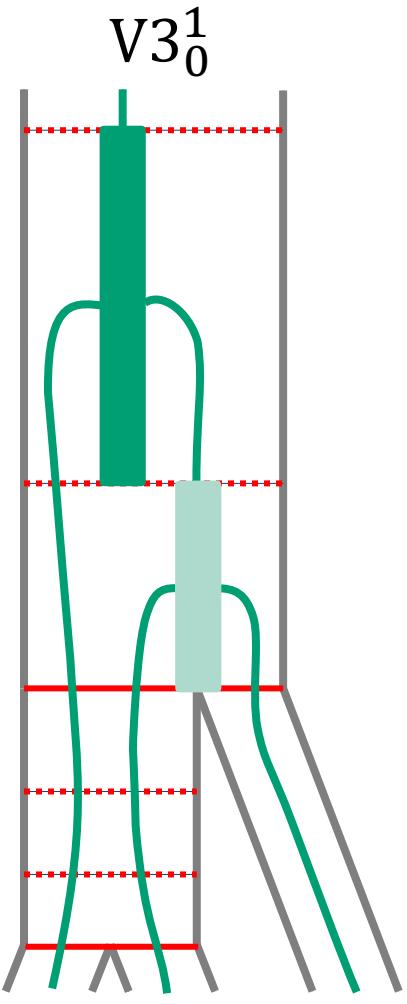
Simulating from the model



Simulating from the model



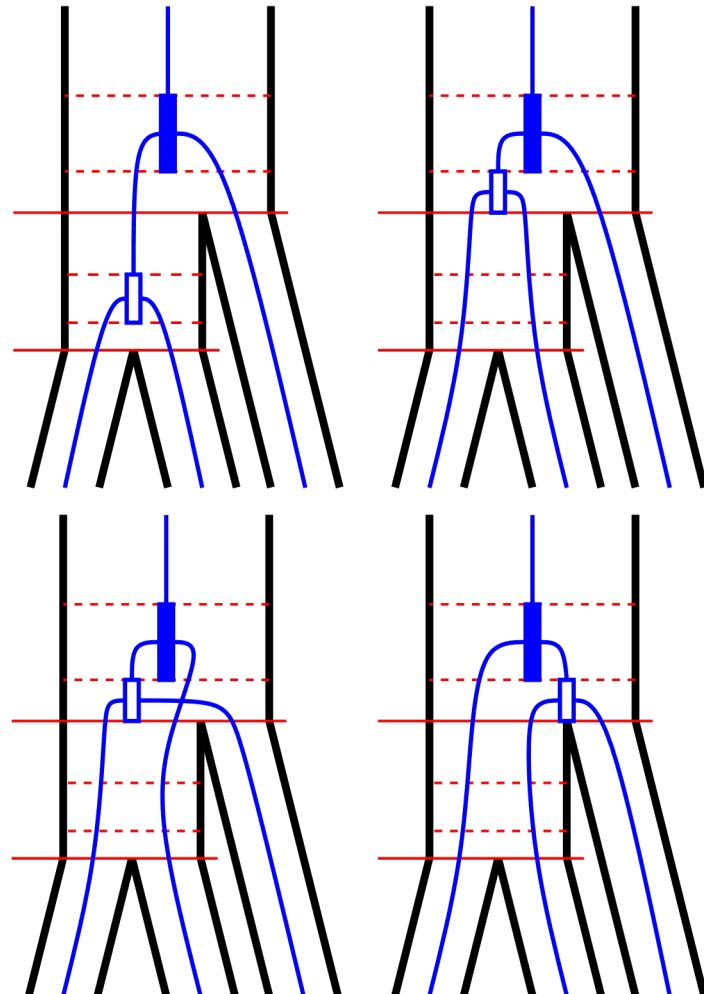
Simulating from the model



Discussion

Current state:

- Transition probability matrix calculated



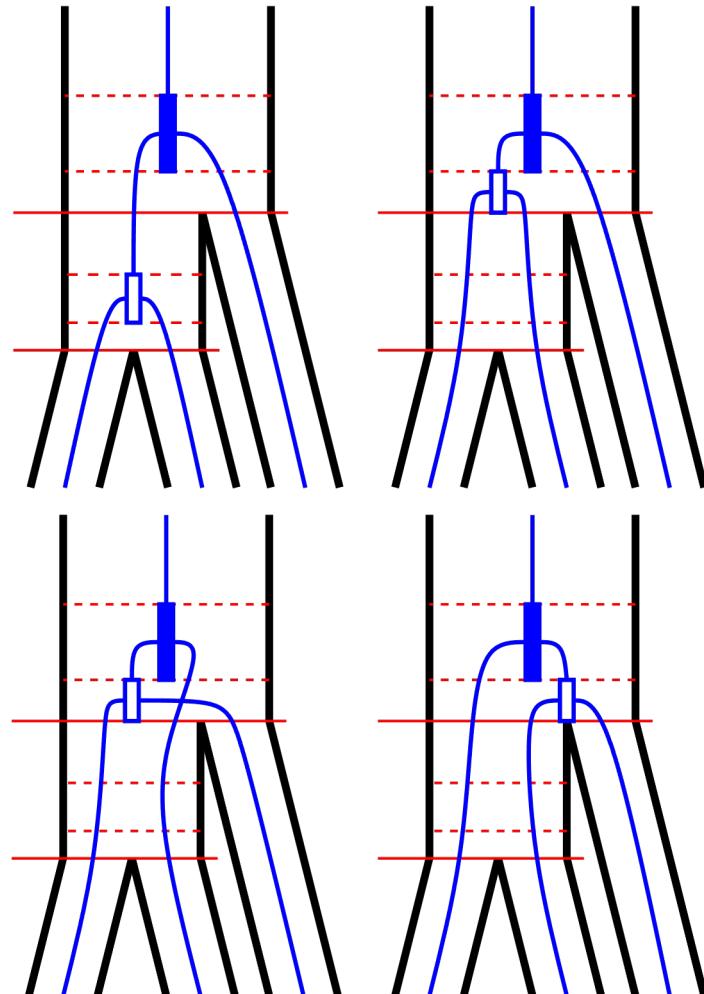
Discussion

Current state:

- Transition probability matrix calculated

Future work:

- Optimize transition probability matrix
- Emission probabilities.
- Choose and code optimization procedure
- Analyze simulated data



Discussion

Current state:

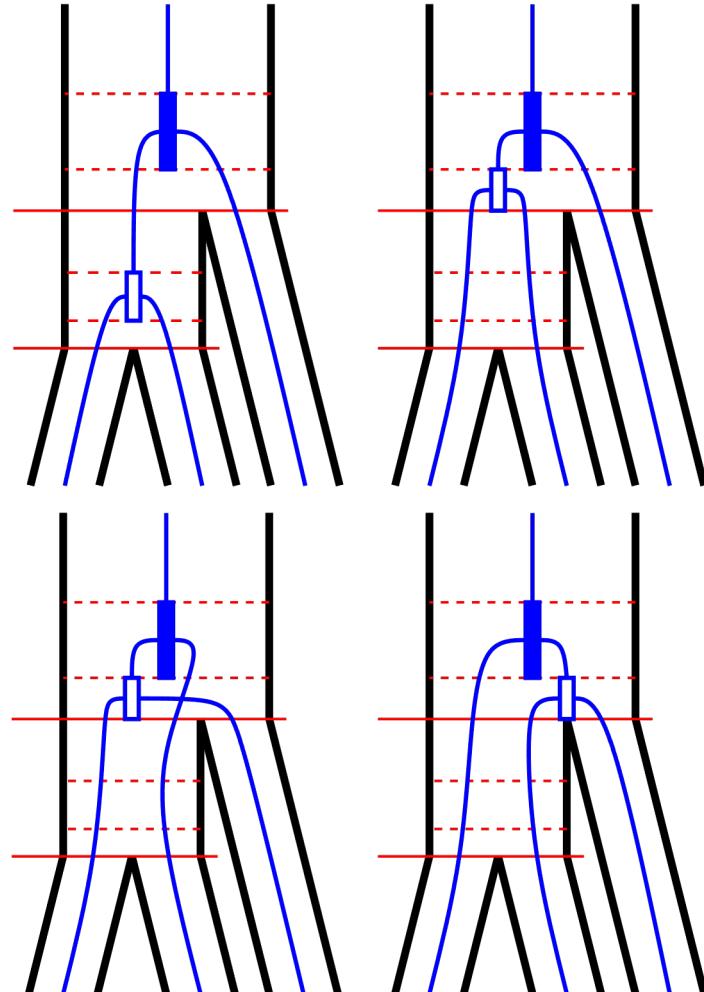
- Transition probability matrix calculated

Future work:

- Optimize transition probability matrix
- Emission probabilities.
- Choose and code optimization procedure
- Analyze simulated data

Future perspective:

- ILSMC is flexible:
 - Different parameterization per interval
- ILSMC is extensible:
 - ILS and introgression



Acknowledgements



Mikkel H. Schierup



Asger Hobolth

Julien Y. Dutheil



My colleagues at BiRC

References

- Dutheil, J. Y., Ganapathy, G., Hobolth, A., Mailund, T., Uyenoyama, M. K., & Schierup, M. H. (2009). Ancestral population genomics: the coalescent hidden Markov model approach. *Genetics*, 183(1), 259-274.
- Mailund, T., Dutheil, J. Y., Hobolth, A., Lunter, G., & Schierup, M. H. (2011). Estimating divergence time and ancestral effective population size of Bornean and Sumatran orangutan subspecies using a coalescent hidden Markov model. *PLoS genetics*, 7(3), e1001319.
- Li, H., & Durbin, R. (2011). Inference of human population history from individual whole-genome sequences. *Nature*, 475(7357), 493-496.

Contact



Iker Rivas-González - PhD Student

Bioinformatics Research Centre (BiRC)
Aarhus University

✉ irg@birc.au.dk @irg_bio

t11-rivas-gonzález-phylogenetically



AARHUS
UNIVERSITY