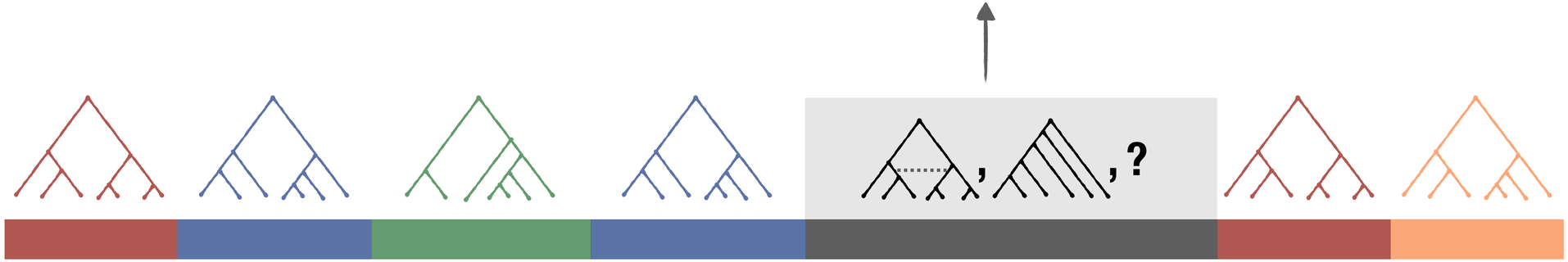
Mapping Strongly Discordant Regions on the Genome Using Hidden Markov Models

Ali Osman Berk Şapcı, Shayesteh Arasti, and Siavash Mirarab
UC San Diego

Coalescence with recombination versus non-ILS discordance

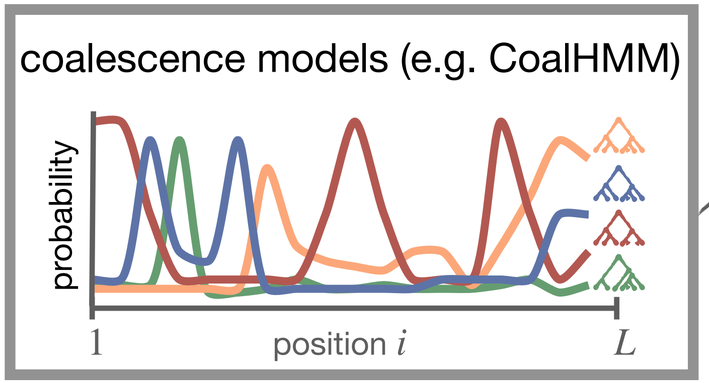
Expected discordance:
incomplete lineage sorting (ILS) & recombination

Strong non-ILS discordance:
biological (e.g., hybridization) & artifacts (e.g., errors)



Chromosome

Can we detect outlier regions using the posterior probabilities across the loci?



Expensive to compute!

	1	2	…	i	…	L-1	L
	0.03	0.03	…	0.09	…	0.26	0.22
	0.04	0.05	…	0.01	…	0.11	0.12
	…	…	…	…	…	…	…
	0.23	0.24	…	0.11	…	0.09	0.10
	0.06	0.06	…	0.32	…	0.03	0.04

posterior probabilities

0.22
0.12
…
0.10
0.04

stationary distribution

An HMM w/ locus tree summary statistics as emissions

Given a species tree and gene trees sampled across a sequence:
- QQS values as emissions—*fast*
- approximating the emission distribution via simulations
- informative priors on transitions to deal with noise