

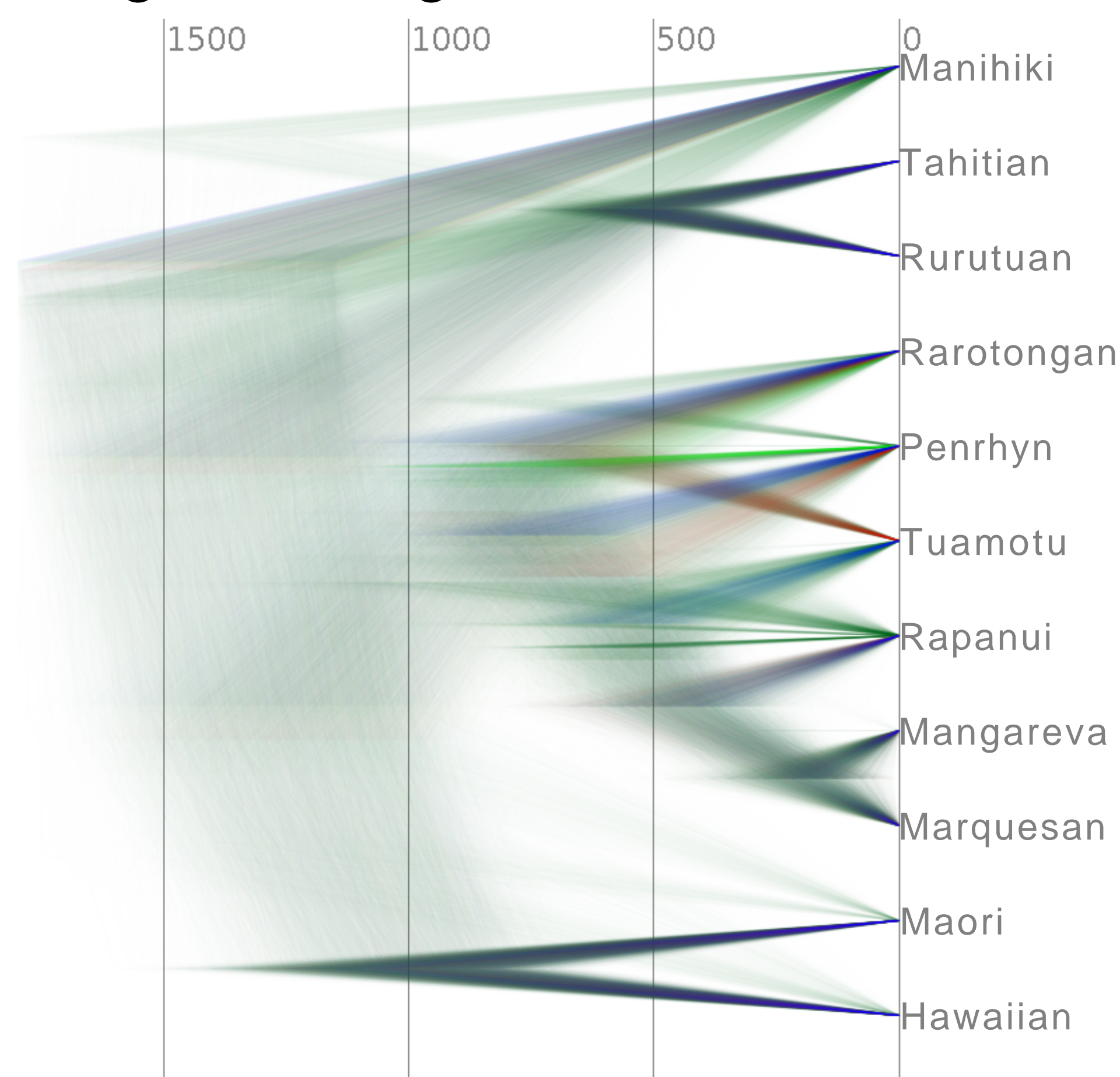
Lagged couplings for phylogenetic inference

Diagnosing convergence of MCMC on trees

Luke J. Kelly, Robin J. Ryder and Grégoire Clarté

Problem

Phylogenetics is the problem of reconstructing the ancestral history of a set of taxa descended from a common ancestor. MCMC is the primary tool for Bayesian phylogenetic inference. **We lack methods to properly quantify MCMC convergence or mixing on the space of trees and model parameters** so struggle to separate modelling and fitting errors.



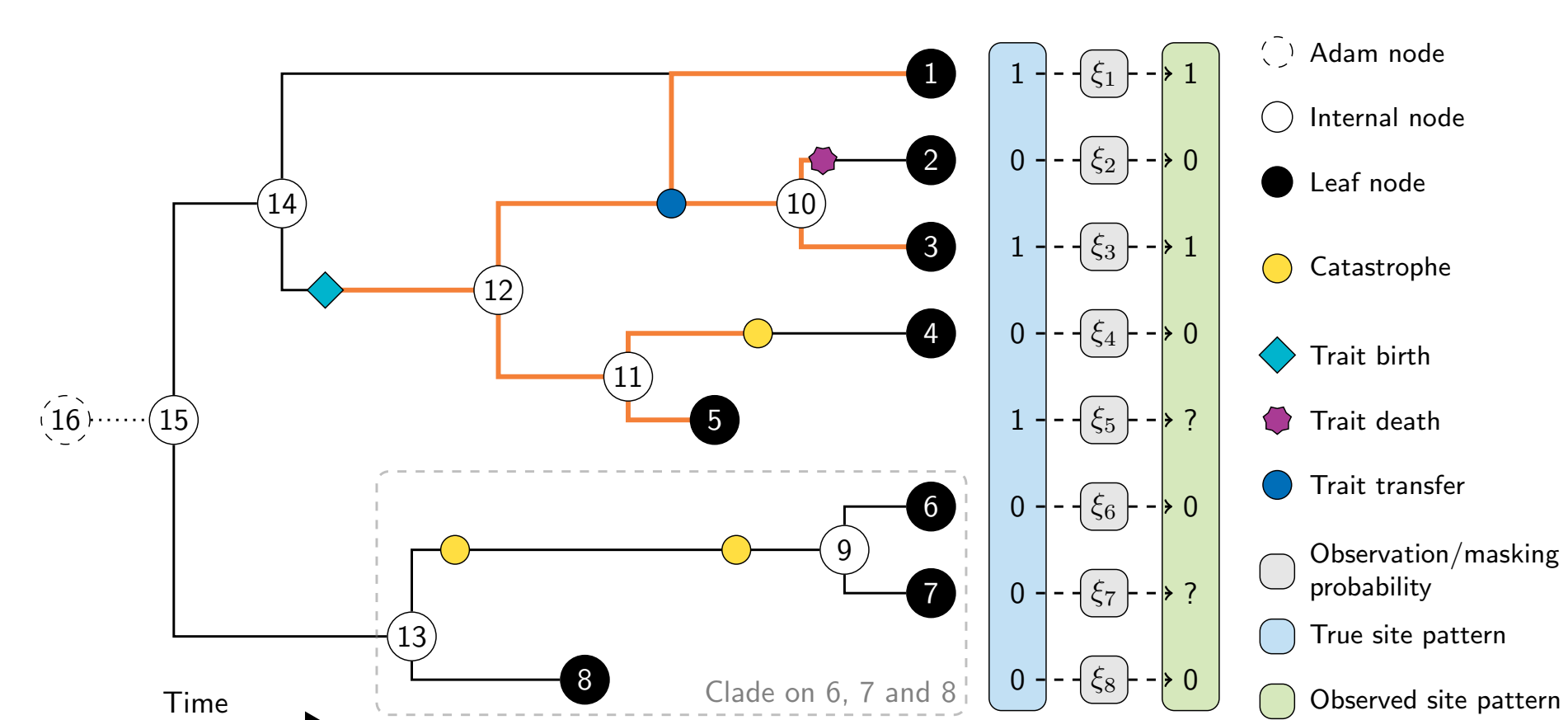
MCMC samples for the tree component (topology and branch lengths) of a phylogenetic posterior distribution; data are lexical traits in Eastern Polynesian languages, time is in years before the present, and the most frequently sampled topology is in blue.

Stochastic Dollo model

A branching process \circ on sets of traits is the phylogeny of the observed taxa \bullet .

- New traits are **born** \blacklozenge at rate λ .
- Trait instances **die** \blackstar independently at rate μ and **transfer** \bullet copies to other species at per capita rate β .

We record the terminal outcome of this process at the leaves.



A phylogenetic tree and **trait history** drawn from the Stochastic Dollo model. **Catastrophes** \bullet represent bursts of evolutionary activity; data are missing at random at each leaf.

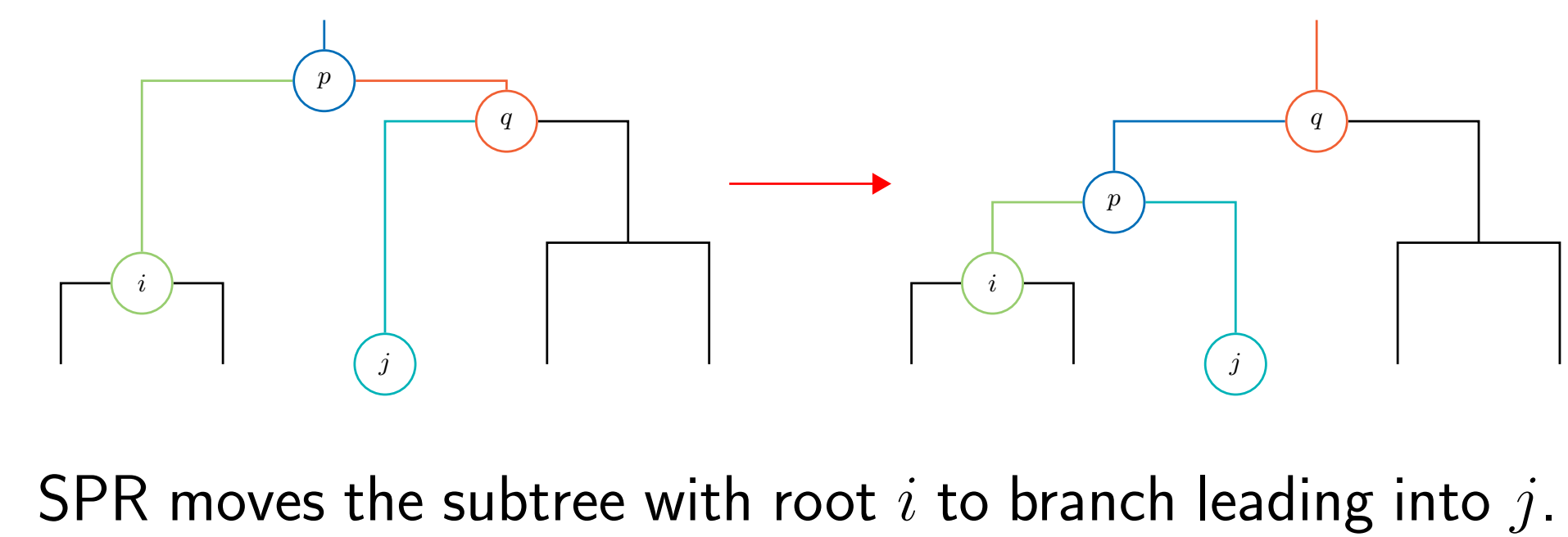
We want to infer the tree, model parameters and catastrophe locations.

References

- P.E. Jacob, J. O'Leary, and Y.F. Atchadé. *J. Roy. Statist. Soc. B*, 82(3):543–600, 2020.
N. Biswas, P.E. Jacob, and P. Vanetti. *NeurIPS*, pages 7389–7399, 2019.
L.J. Kelly, R.J. Ryder, and G. Clarté. *arXiv 2108.13328*, 2021.

MCMC on phylogenies

Our MCMC proposal kernel is a mixture of 19 local kernels. Subtree prune-and-regraft (SPR) proposals detach a randomly chosen subtree and reattach it at a randomly chosen destination along the tree.



Existing methods to diagnose MCMC convergence on trees are *ad hoc*, such as comparing multiple low-dimensional summaries across one or more chains.

Coupling

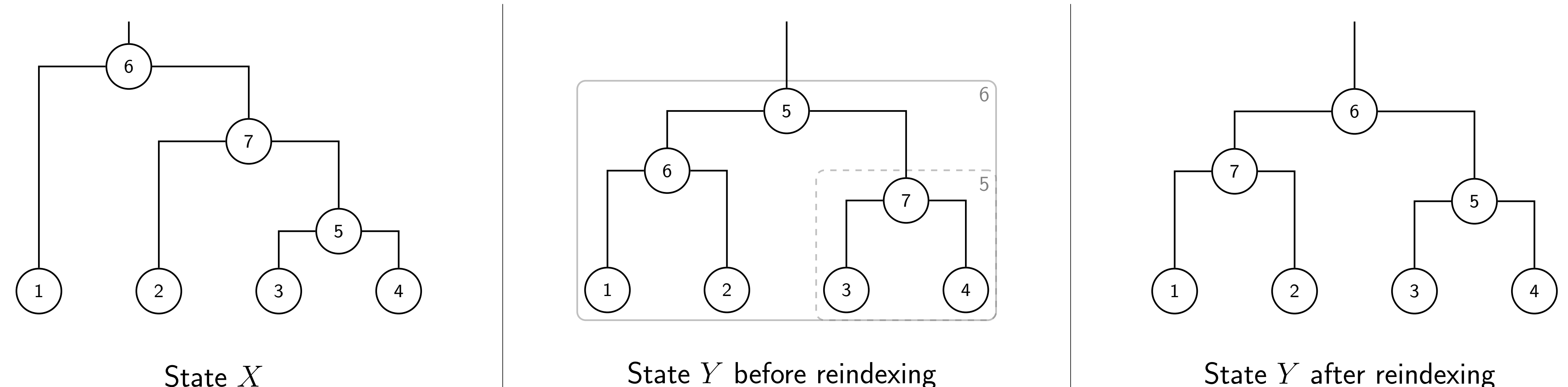
A coupling of two distributions p and q on space \mathcal{X} is a joint distribution γ with marginals p and q . In order to debias MCMC estimators, Jacob et al. (2020) describe how to couple the MCMC transition kernels for lag-1 staggered chains (X_t) and (Y_{t-1}) with common target π so that they

meet at a finite random time τ and remain faithful thereafter. Biswas et al. (2019) use chains coupled at lag l to estimate the following bound on the total variation distance between $\pi_t = \mathcal{L}_{X_t}$ and π ,

$$d_{TV}(\pi_t, \pi) \leq \mathbb{E} \left[0 \vee \left\lceil \frac{\tau - l - t}{l} \right\rceil \right].$$

Coupling tree proposals

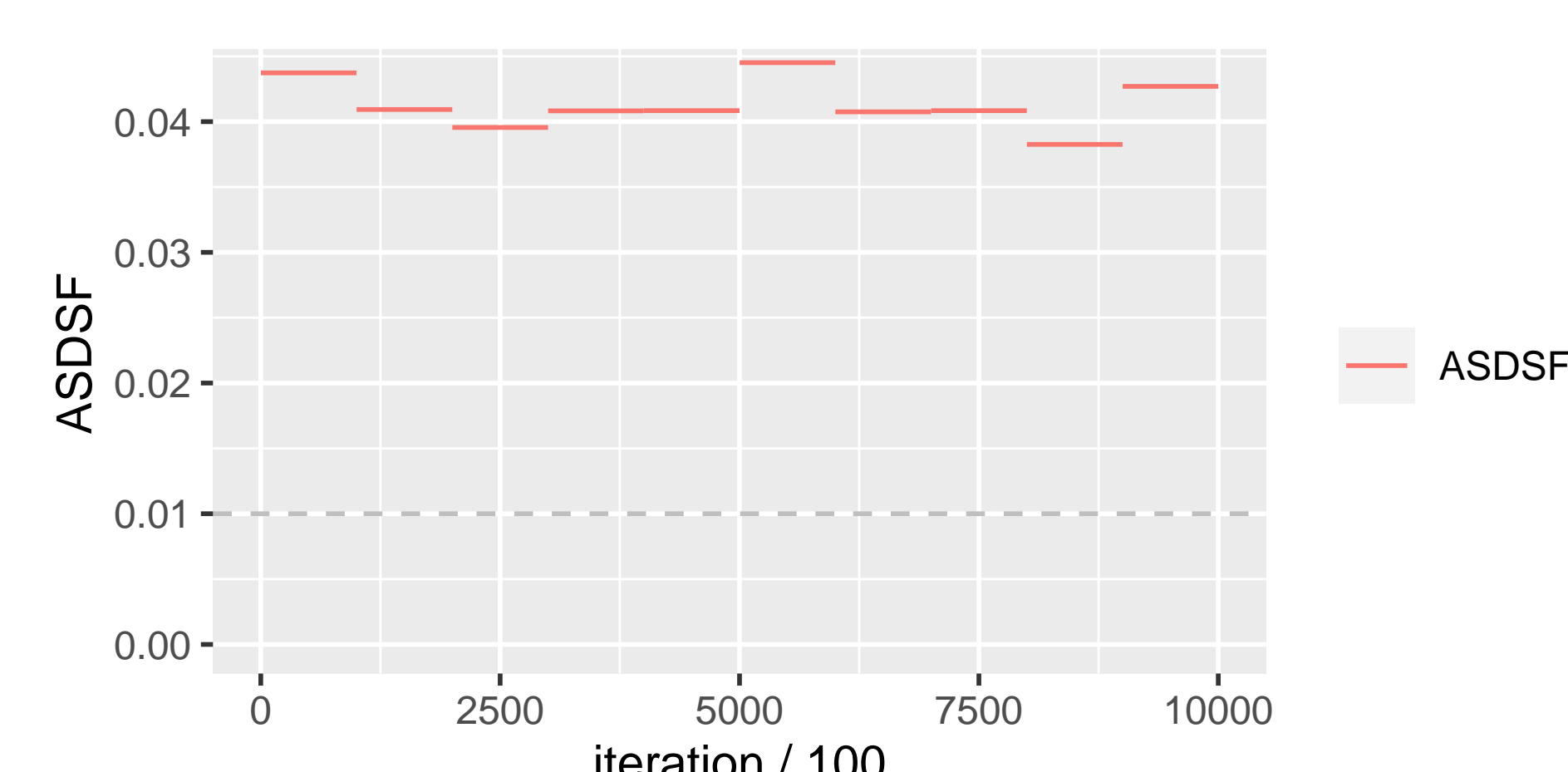
At each iteration, we reindex Y so that nodes with matching leaf sets have identical indices.



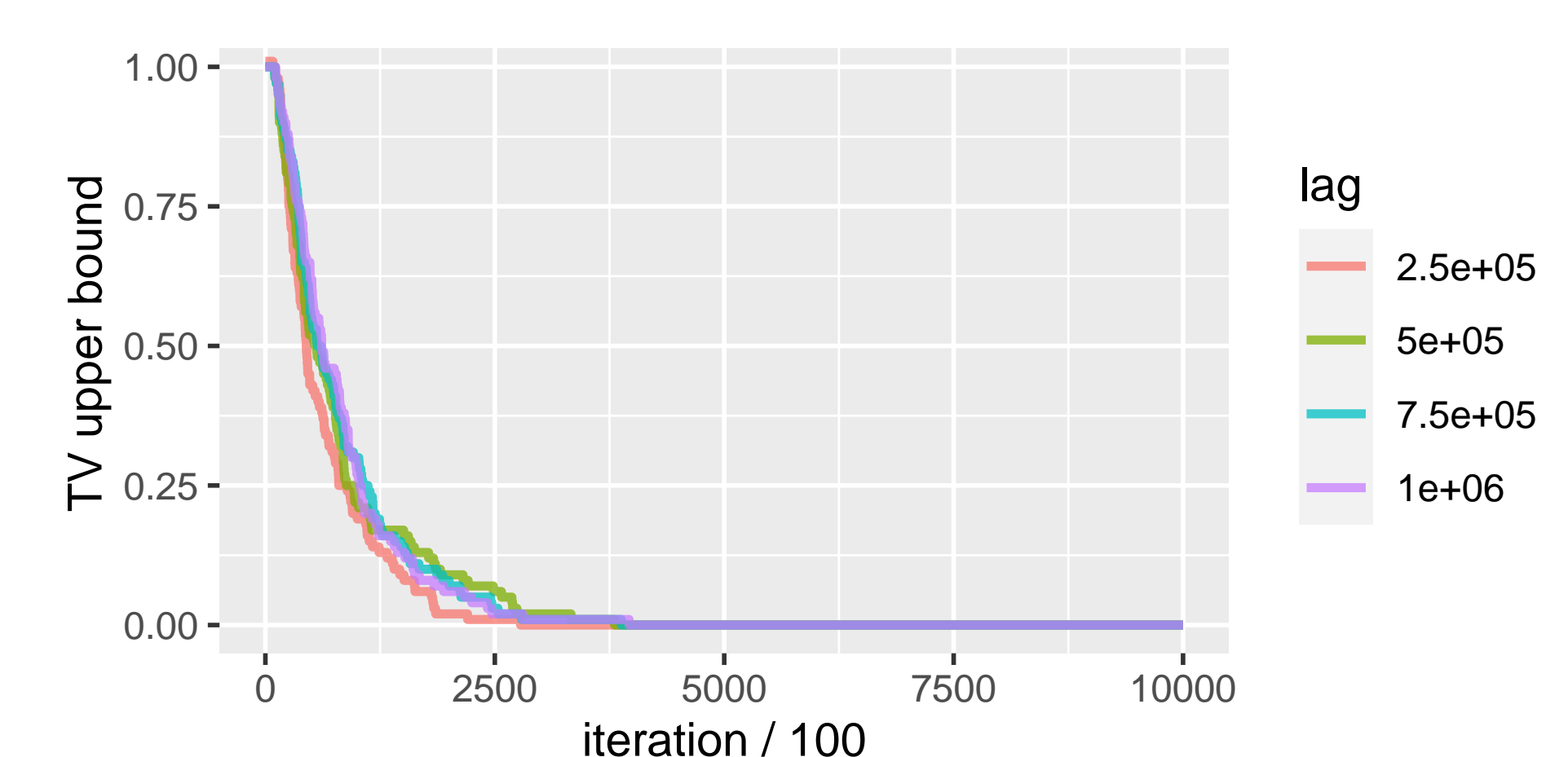
We then sample from a coupling at each step of a proposal to bring the states closer together. In the above example, if we propose to move subtree with root $i = 1$ in both and sample the destination branches from a maximal coupling, then the proposed topologies will be identical.

Illustration

We fit the Stochastic Dollo model to lexical trait data from Eastern Polynesian languages. We compute the Average Standard Deviation of Split Frequencies (ASDSF) on sliding windows of samples from 100 marginal chains. We estimate the above bound on the total variation distance between π_t and π from the meeting times of 100 pairs of chains at each of 4 lags.



ASDSF on disjoint sliding windows of samples.



Estimated TV bounds on tree and model parameters.

In contrast to ASDSF and other approaches, **the estimated TV bounds diagnose convergence on the entire space** of trees, branch lengths and other parameters. Samples from coupled chains can also be used to construct unbiased estimators.