

Inferring long-range gene-flow events in a background of IBD with FEEMS

Vivaswat Shastry¹ and John Novembre^{1,2}

¹Committee on Genetics, Genomics & Systems Biology,

²Department of Human Genetics,

University of Chicago, Chicago IL

vivaswat@uchicago.edu



Introduction

- Long-range gene-flow events are not well modeled by continuous migration surfaces
- Such perturbations can be modeled as recent admixture events over a background equilibrium migration surface inferred using isolation-by-distance principles

Methods

- We model the **pairwise coalescent time** between outlier demes as a function of (instantaneous) admixture proportion and neighboring migration rates
- We test our method on data simulated using *msprime* (Kelleher *et al*, 2016) and apply it a dataset of the greenish warblers ring species complex in Asia (Alcaide *et al*, 2014)

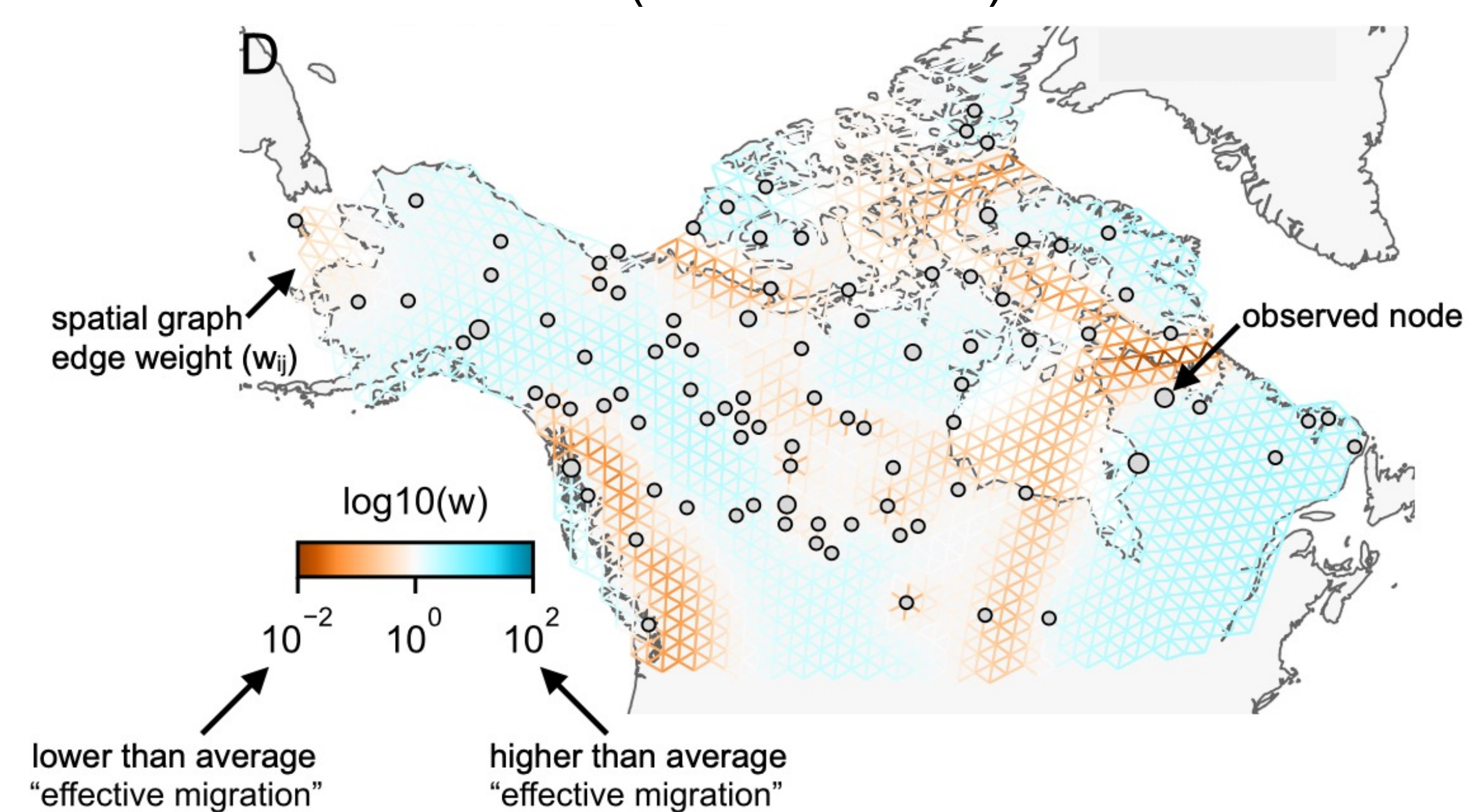
Results

- In simulations, we find high accuracy (>80% for $c > 0.25$) in detecting edges & infer underlying admixture proportions with low bias & low variance (**RMSE ≈ 0.05**)
- In the warblers dataset, we infer a long-range admixture event that was previously uncovered by *SpaceMix* (Bradburd *et al*, 2018) **plus** a few new events that map to historical patterns of genome sharing in the species

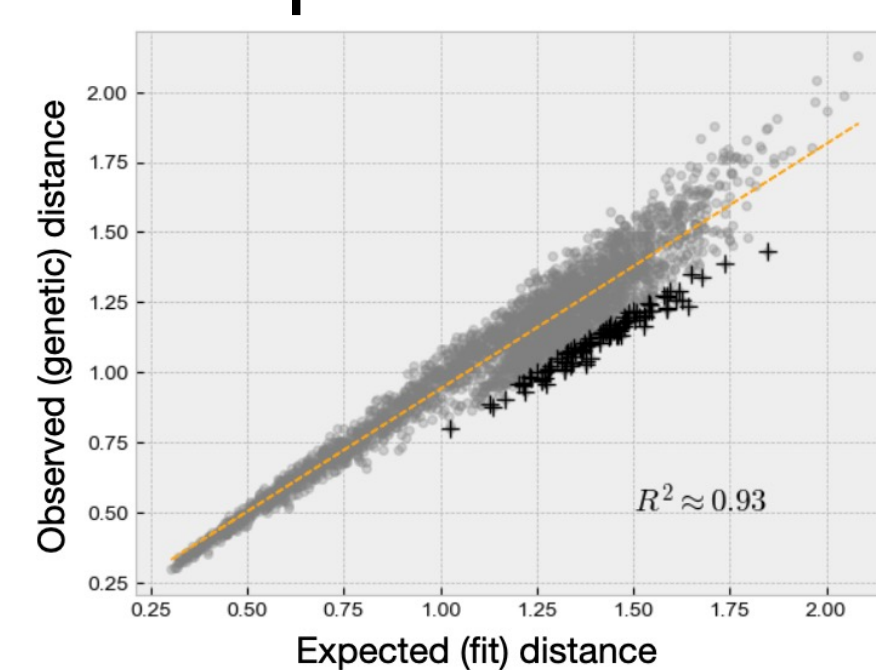
Assumption:

- **Stationary, symmetric migration surfaces**

Baseline model (arctic wolves)



1. Detect pairs of outlier demes



Demes that have a **smaller** observed (genetic) distance than expected (fit) could be due to an unmodeled long-range admixture component

Repeat

2. Model pairwise coalescent times as a function of admixture proportion c

$$T'_{dd} = (1 - c^2)T_{dd} + 2c(1 - c)T_{sd} + c^2T_{ss}$$

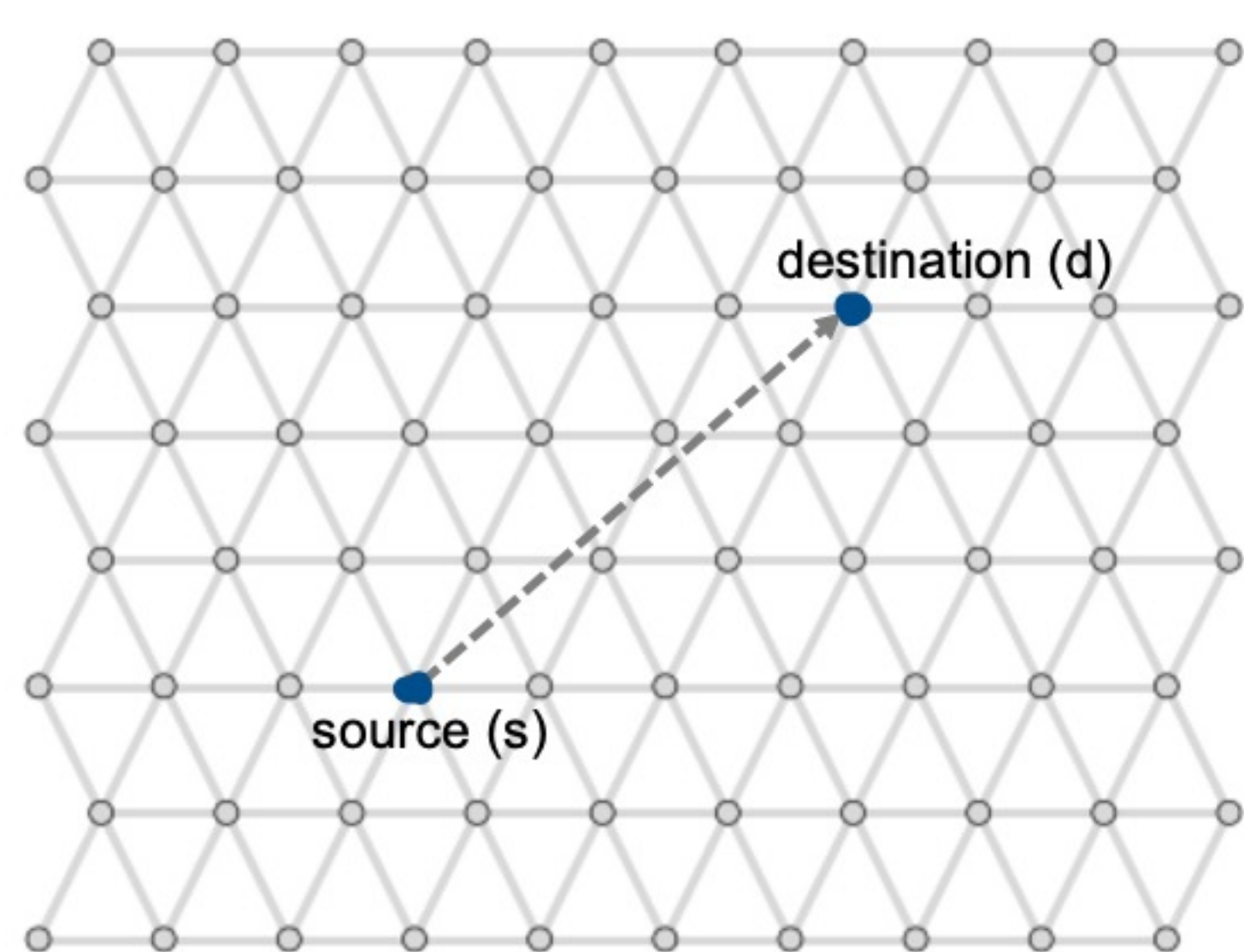
$$T'_{sd} = cT_{ss} + (1 - c)T_{sd}$$

$$T'_{ss} = T_{ss}$$

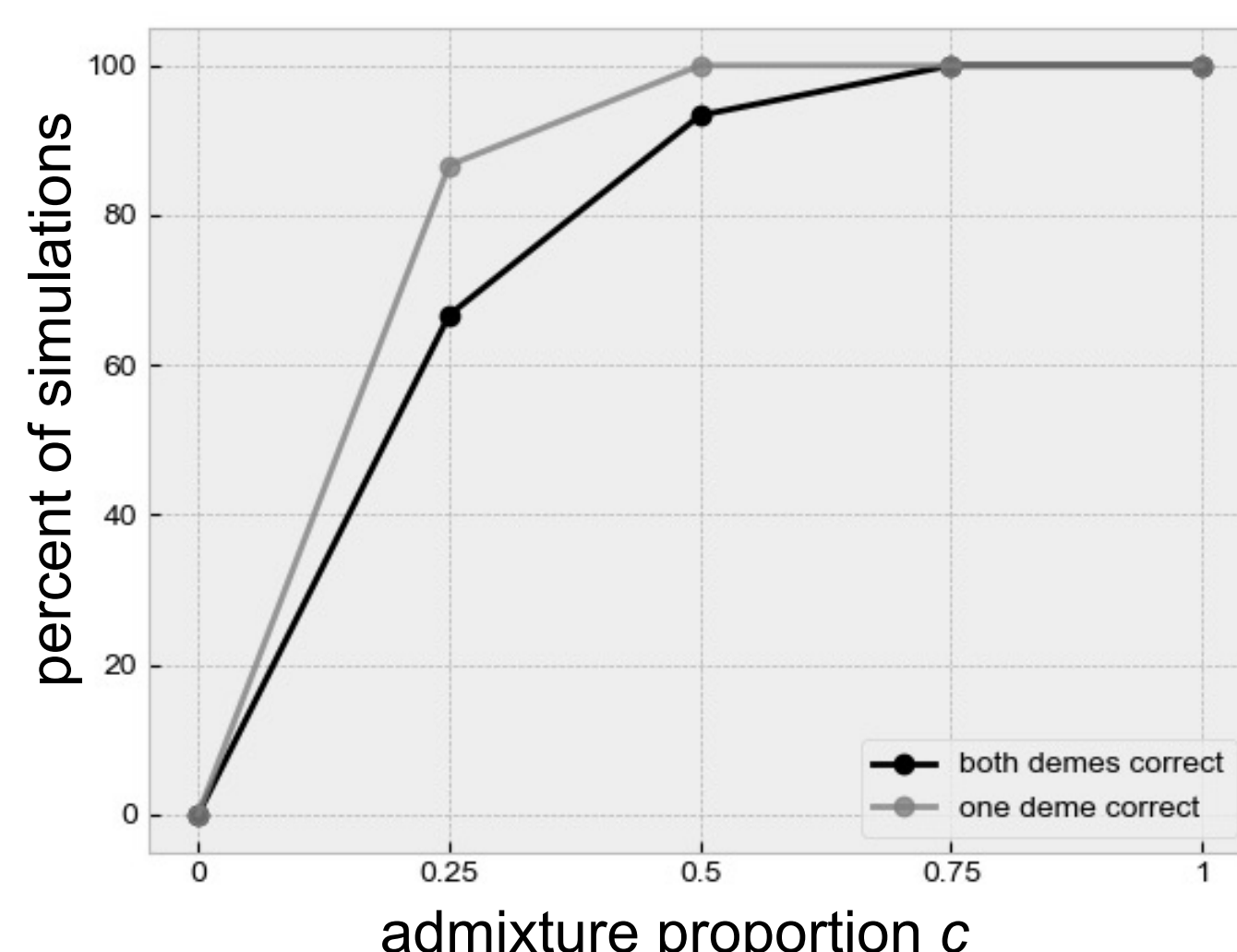
3. Choose edge with highest joint log-likelihood

Results from *msprime* simulations with a simple grid

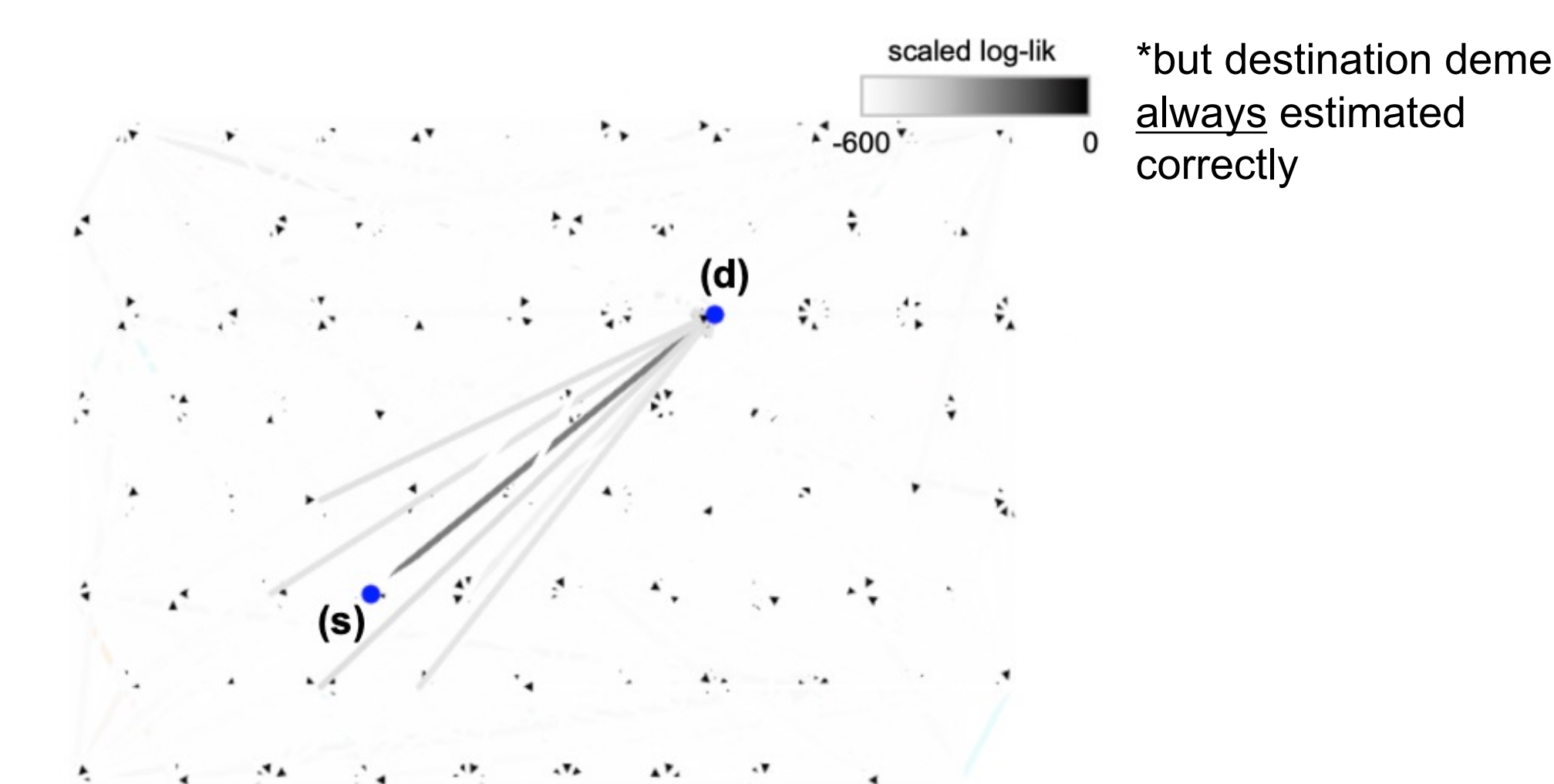
Setup: 8 x 10 grid of 15 diploid inds/deme and 1,500 unlinked SNPs with instantaneous admixture event



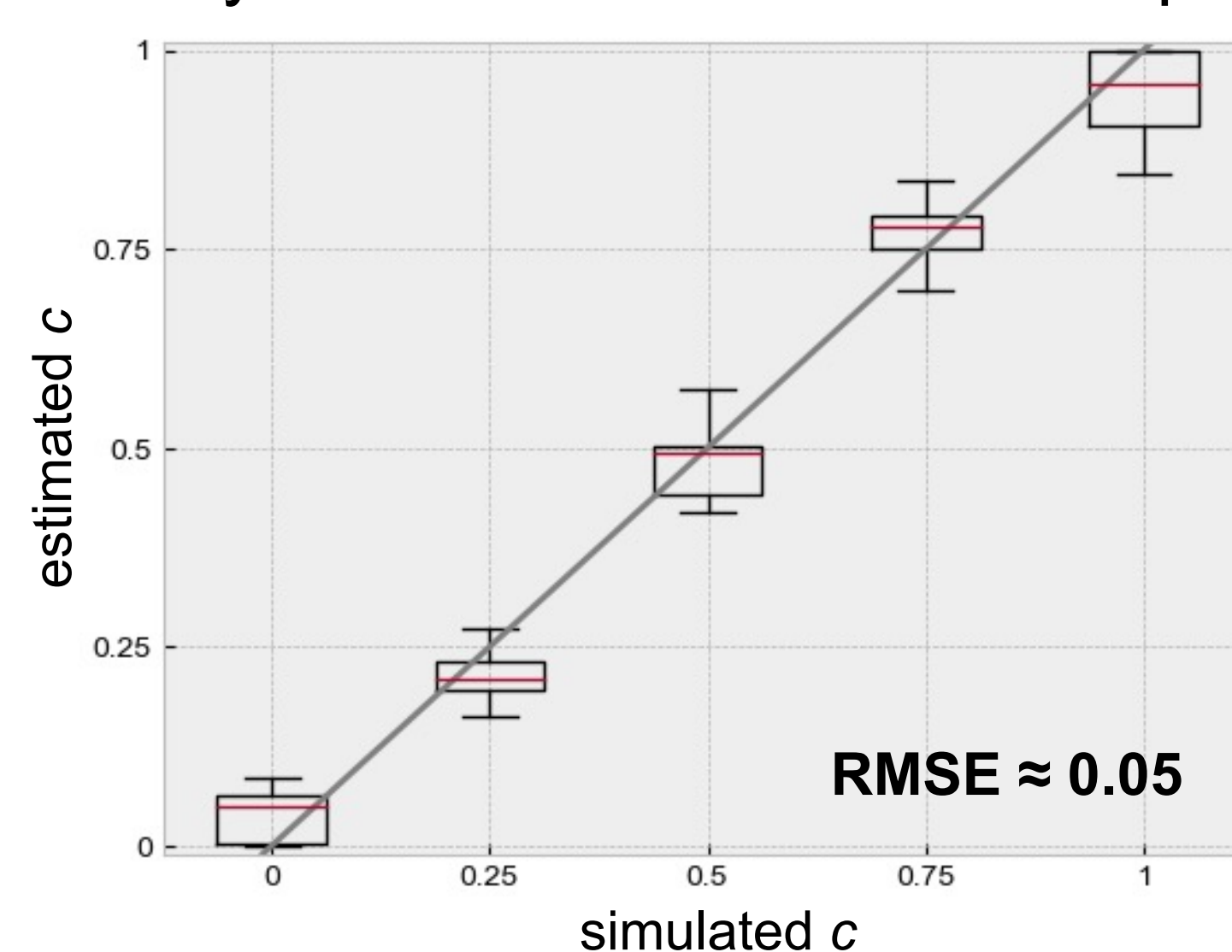
1. Recovers true edge over 80% of time with $c > 0.25$



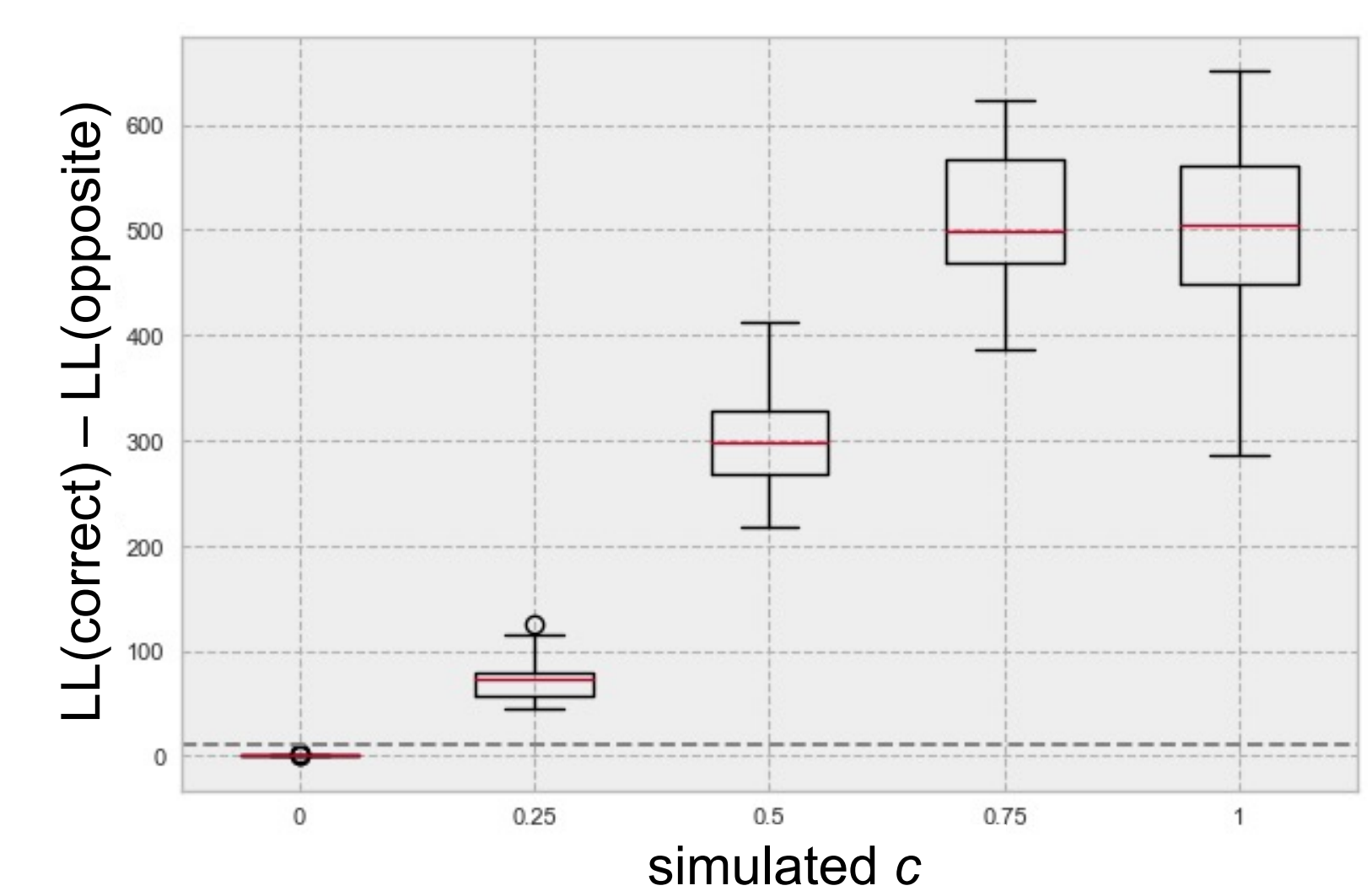
2. Potential source deme estimated with some error*



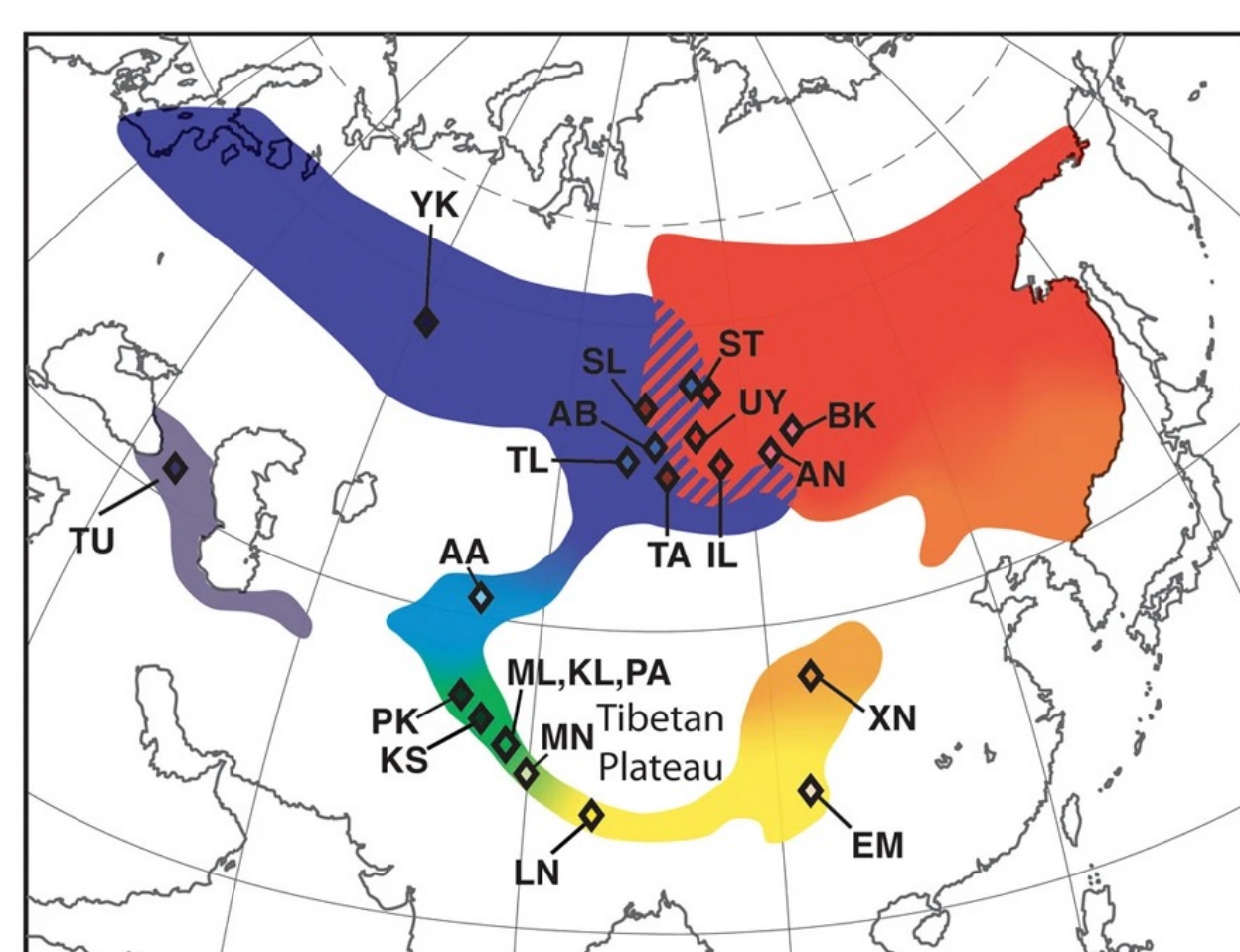
3. High accuracy of estimation for admixture proportions



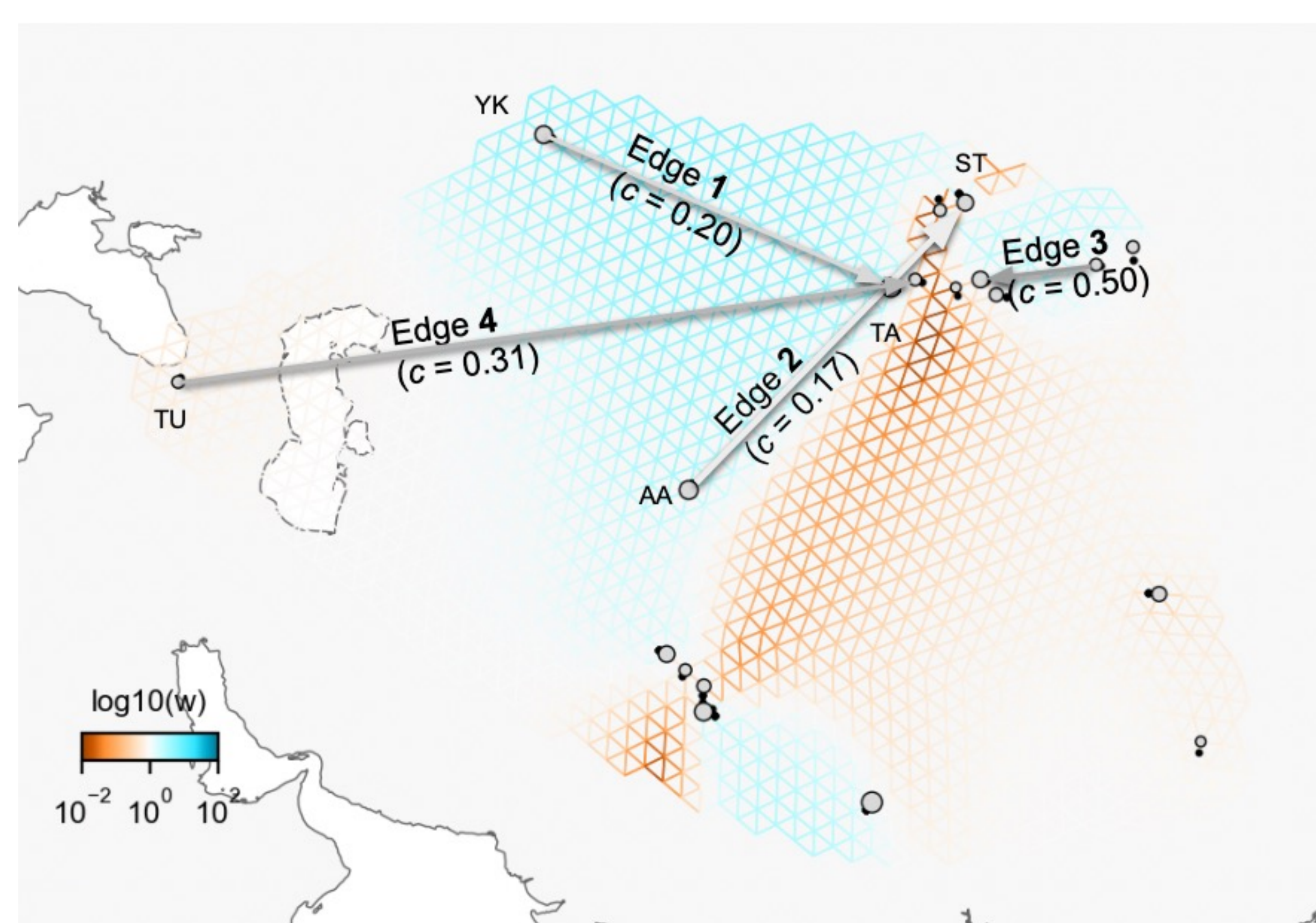
4. Detects **correct** direction of admixture



Results from *Phylloscopus trochiloides* species complex (Alcaide *et al*, 2014)



Setup: 95 inds across 19 demes and ~2,250 SNPs



Conclusions

Edges 1 & 3 track migrations from outer populations to the hybrid zone – indicating ancient admixture events (*new*)

Edge 2 captures the same long-range admixture event (both direction & magnitude) as *SpaceMix* ($c = 0.19$, 95% CI: 0.15 – 0.24)

Edge 4 finds an event from a deme in Turkey to the Russian hybrid zone – genome sharing between outer demes & central Asian demes (*new*)

Future directions

1. Incorporate older admixture events with $\tau > 0$
2. Allow sources to be picked from **unsampled** demes (à la *SpaceMix*)
3. Implement method to assess **significance** of admixture events

References

- Alcaide, et al. *Nature* (2014) [greenish warblers data set]
- Bradburd, et al. *PLoS Genetics* (2016) [*SpaceMix*]
- Kelleher et al. *PLoS Computational Biology* (2016) [*msprime*]
- Marcus, Ha, et al. *eLife* (2021) [FEEMS]