The Effect of Introgression on the Joint Distribution of Gene Trees at Two Linked Loci

Michael Miyagi, Andrew J. Blumberg, John Wakeley | m_miyagi@g.harvard.edu

We model the history of linked loci sampled from three species with a pulse of gene flow

- We describe the resulting joint gene genealogies for two loci sampled from three species based on the work of Slatkin and Pollack, adding a one-way introgression event from species B to C.
- We track two loci backwards in time, between which recombination occurs with probability r per generation.
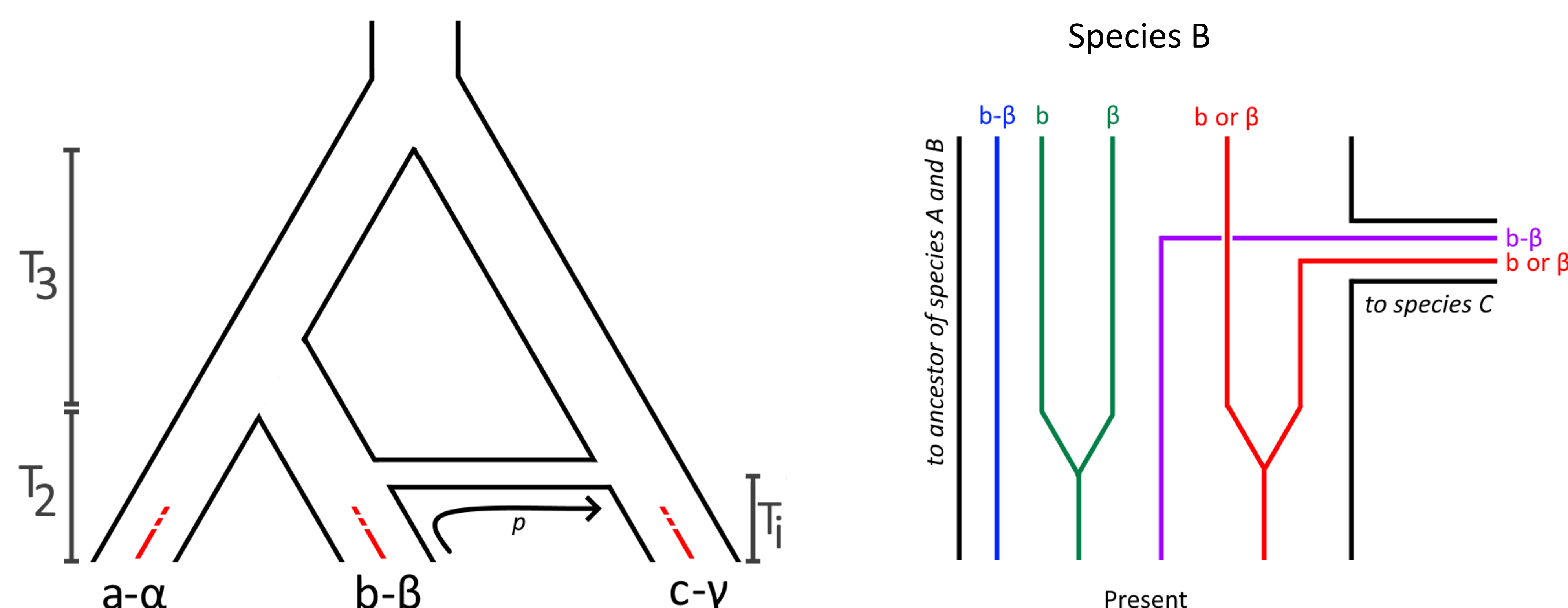

Figure 1. We sample one individual from each species in the present day. We allow the loci of interest to recombine off the same chromosome and keep track of this in addition to potential coalescence events.

Figure 2. Some of the possible states for the lineage in species B. If it does not introgress, it joins species A with both loci on a single chromosome (blue) or two separate ones (green). If introgression occurs, it can be at only one locus (red), at both on the same chromosome (purple), or at both independently (not shown).

Total concordance probability can be non-monotonic with introgression fraction

- The probability that the two loci have the same topology depends on both the recombination distance and introgression probability.
- The concordance probability is minimized when the probability that the two loci take different paths in the species network is maximized.
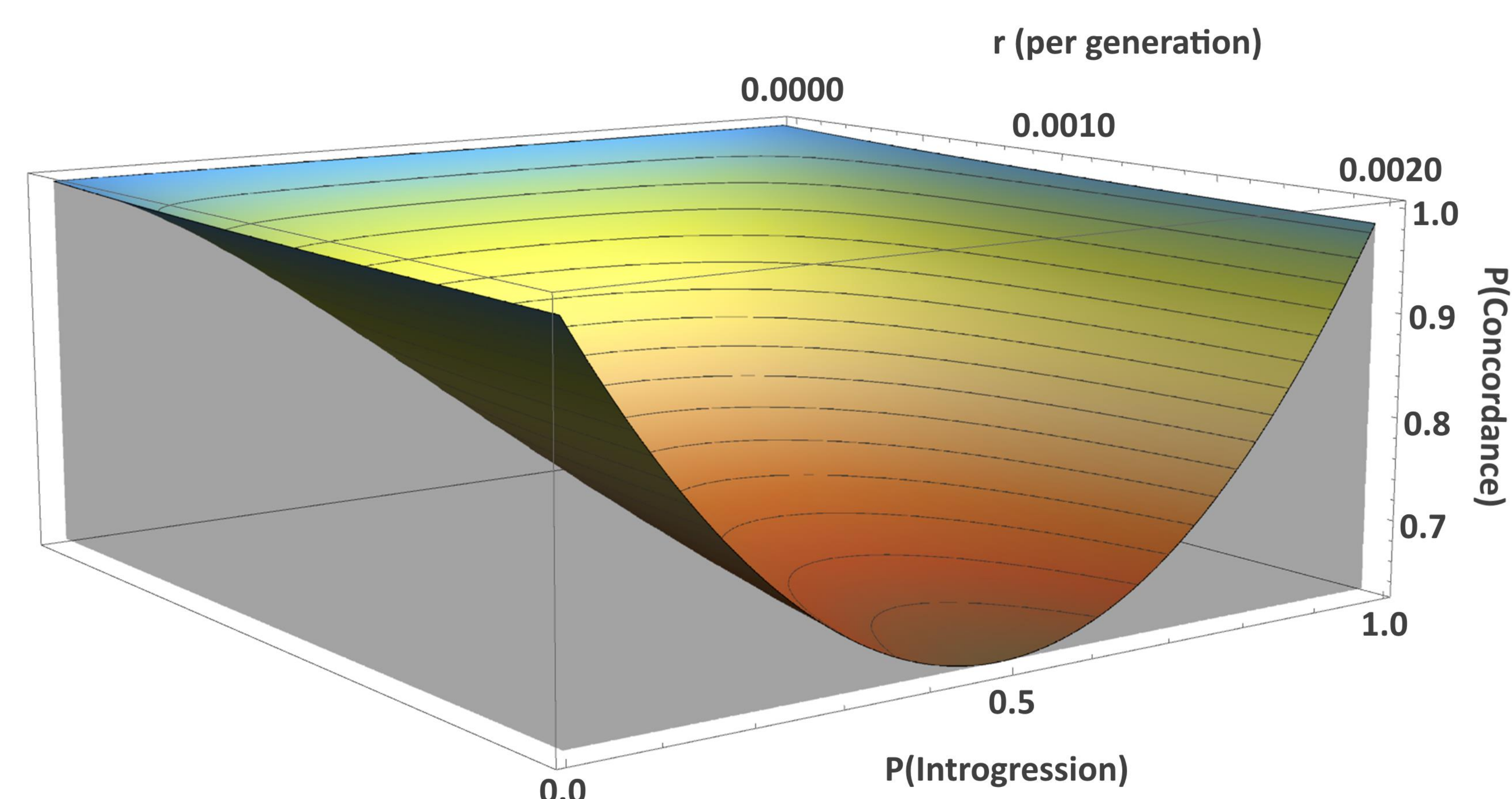

Figure 3. The probability that the two sampled loci have matching gene tree topologies as a function of the recombination distance between them and the probability of introgression. Results for constant $N=2000$ and $T_3=T_2=2*T_i=2000$ generations.

Even conditioned on deep coalescence, the joint coalescence times are not independent of topology

- While the coalescence time at one locus conditioned on incomplete lineage sorting (ILS) is independent of its topology, the expected time to the first coalescence event from a pair of loci changes with its identity.
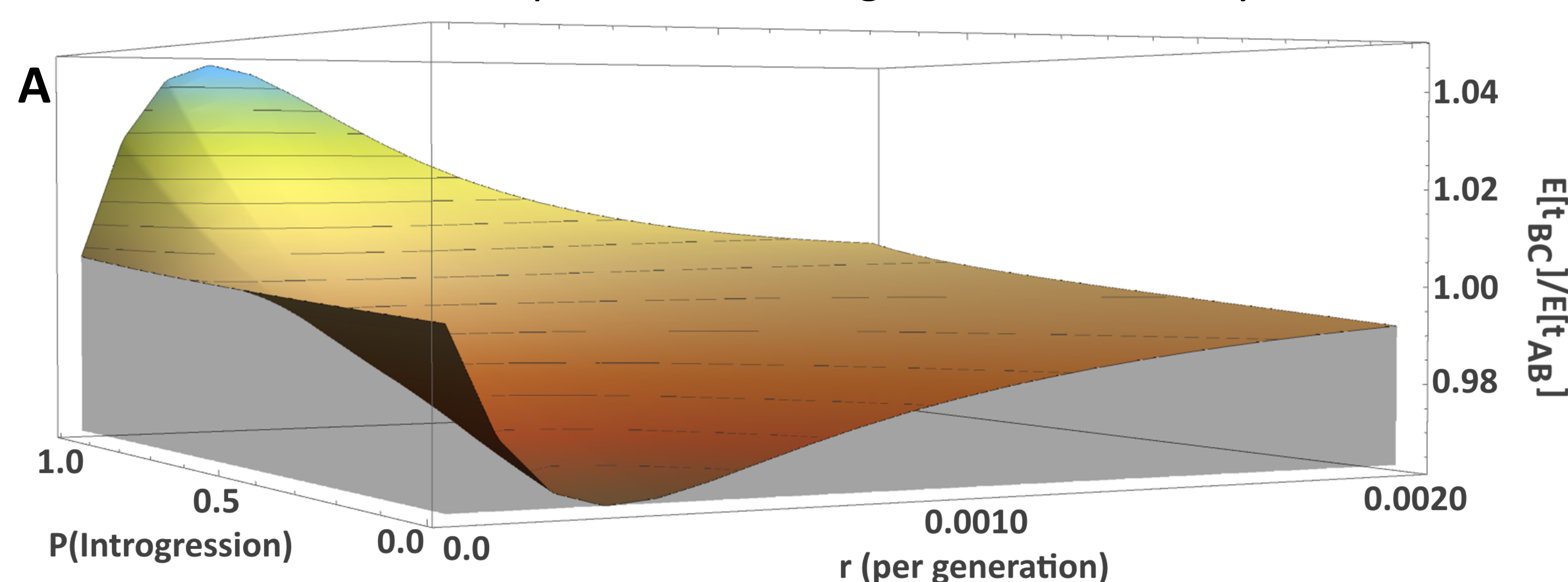

Figure 4. The ratio of ILS coalescence times conditioned on their topology. t_{XY} refers to the time when species X and Y coalesce conditioned on ILS at both loci. Since the surface is not flat, the timing and topology are not independent, even though ILS has occurred. The species topology is compared to the introgression (**A**) and ILS-only (**B**) ones. Results for constant $N=2000$ and $T_3=T_2=2*T_i=2000$ generations.
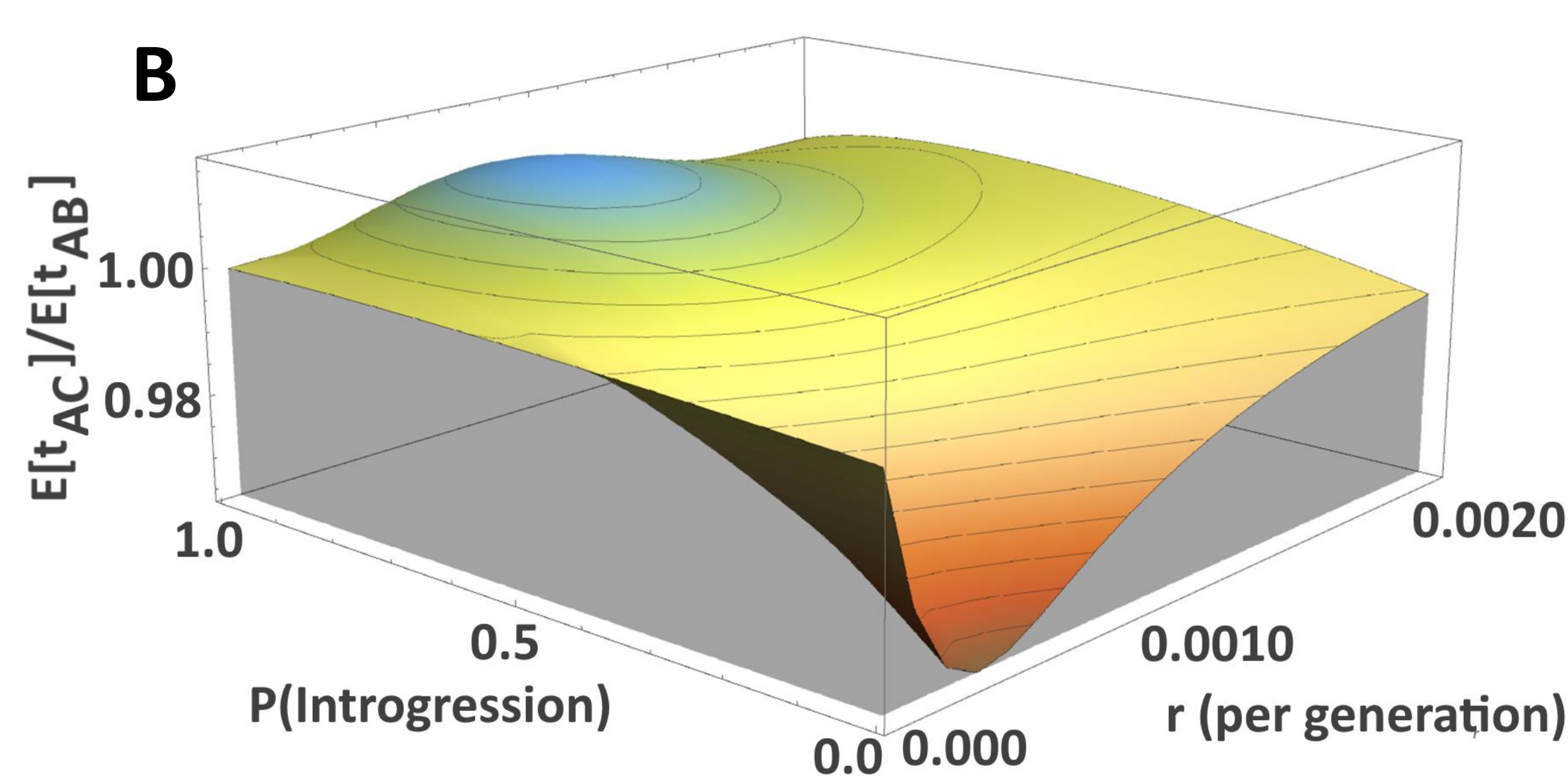
Recombination rate does not uniformly increase the frequency of discordant gene-tree pairings

- Discordant gene tree pairings are all contingent on recombination events, but where these events occur on the tree is also important.
- The magnitude and timing of the introgression event creates intermediate optimal recombination rates for generating certain pairings.
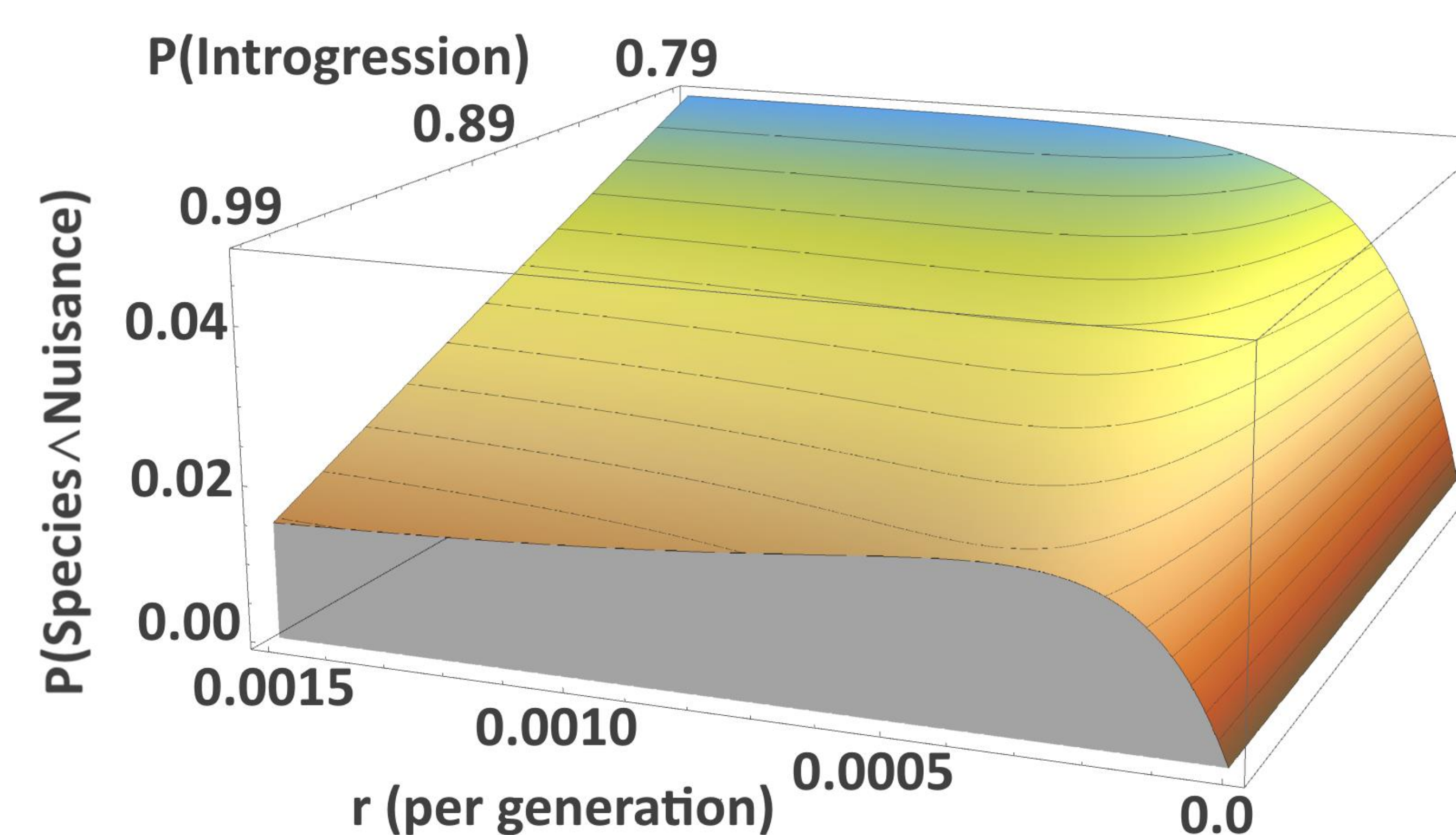

Figure 5. The probability of recovering one species topology and one nuisance (ILS-only) topology. At high introgression probabilities, if recombination is too frequent, the tracked loci are likely to split onto different individuals and increase the likelihood of a coalescent event during the introgression branch. Results for constant $N=2000$ and $T_3=T_2=2*T_i=2000$ generations.

References
Slatkin and Pollack 2006. The Concordance of Gene Trees and Species Trees at Two Linked Loci. *Genetics*.
Simonsen and Churchill 1997. A Markov chain model of coalescence with recombination. *Theor. Popul. Biol.*