**Pairwise likelihoods**
- Recombination maps often estimated by combining likelihoods at pairs of SNPs within a composite likelihood:
  \[ L(\rho) = \prod_{a \neq b} P(n_{ab}; \rho_{ab}) \]
- Previous methods to compute \( P(n_{ab}; \rho_{ab}) \) assume constant population size, leading to inaccuracies such as spurious recombination hotspots.
- Our new method **LDpop** computes pairwise likelihoods under changing population size, improving the accuracy of inferred maps.

**An exact formula for \( P(n, \rho) \)**
We compute \( P(n_{ab}; \rho_{ab}) \) by constructing a process \( \{M_t\} \) that contains the 2-locus coalescent embedded within it.
- \( \{M_t\} \) constructed in 2 steps: step A constructed backwards-in-time, step B constructed forwards-in-time.
- \( P(M_0) \) given by a product of sparse matrix exponentials:
  \[ P(M_0 = n, \rho) = \left( \prod_{d=1}^{D} e^{\lambda_d} \right) v \]
  where \( D \) is the number of population size changes.

**A fast approximation \( \hat{P}(n, \rho) \)**
- We also develop a fast approximation based on a Moran model with \( N \geq n \) particles.
- Computed with a similar product of sparse matrix exponentials,
  \[ \hat{P}(n, \rho)_n = \left( \prod_{d=1}^{D} e^{\tilde{\lambda}_d} \right) v \]
  but the matrices \( \tilde{\lambda}_d \) are much smaller.

The error in the approximation disappears as \( N \to \infty \). In practice, the approximation is good even when \( N = n \).

**Simulation study: using LDpop improves accuracy**
- Simulated datasets with \( n = 20 \) haplotypes under demography \( \eta(t) \):
  \[ \eta(t) = \begin{cases} 10^6, & t < 10^5 \text{ gens ago} \\ 10^5, & 1.16 \times 10^5 < t \geq 10^5 \text{ gens ago} \\ 10^3, & t \geq 1.16 \times 10^5 \text{ gens ago} \end{cases} \]
- Inferred \( \hat{\eta} \) by computing lookup tables \( \hat{P}(n, \rho)_{n_0} \) under LDpop and providing them to composite likelihood methods LDhat, LDhelmet.
- Using the true or estimated demography \( \eta(t) \) is more accurate than assuming constant \( \eta \).

**Runtime to compute lookup tables**
- Time to compute \( \hat{P}(n, \rho) \) for \( n = 20 \), exact formula

---

**LDpop: 2-locus likelihoods for recombination map estimation under variable population size**
John Kamm¹, Jeffrey Spence², Jeffrey Chan¹, and Yun S. Song¹,³

¹EECS, ²Computational Biology @ UC Berkeley; ³Math, Biology @ UPenn

Contributed equally

https://github.com/popgenmethods/ldpop

---

This research is supported by NIH grant R01-GM108805, NIH training grant T32-HG000047, and a Packard Fellowship for Science and Engineering.

jkamm, spence.jeffrey, chanjed, songyss@berkeley.edu