Microscale estimation of admixture timing

... and an example stochastic process for estimating it ... poorly.

castedo.com/poster

This poster authored in HTML

Why admixture timing?

Genetic data can provide estimates of the timing of mating between populations.

Macroscale vs microscale

	Macroscale	Microscale
Water-wine illustration	stream of wine entering glass	drops of diluted wine
Population genetics	population migrations	mating & chromosome recombination

Why microscale?

1. **Convergent lines of evidence (consilience)** are impacted by the same microscale events of past mating (e.g. genetic evidence & distinctive archaeological cultures in nearby settlements).
2. **Assortative mating** is common in diverse human populations and not modeled by only population migration.

Microscale quantity to estimate

average lineal admixture time

Why lineal admixture time?

1. quantity grounded in realistic model
2. can estimate higher moments of time distribution
3. can estimate fine-grained conditional distributions (e.g. per chromosome)

Definitions

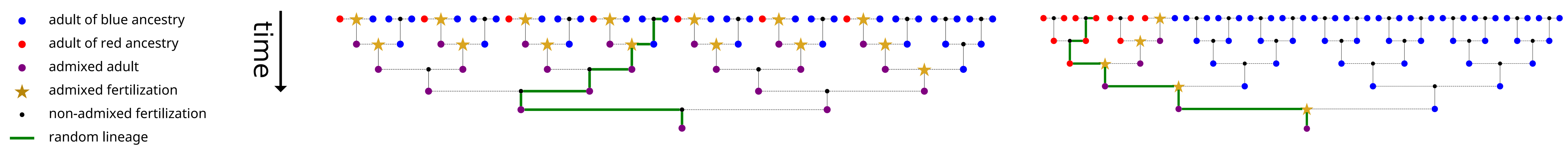
lineal admixture time	the amount of time since fertilization of the first admixed individual in a lineage
lineage	single path of descent in the genealogy of an individual
average lineal admixture time	average across all lineages of all individuals in a population

For more details, visit castedo.com/doc/151.

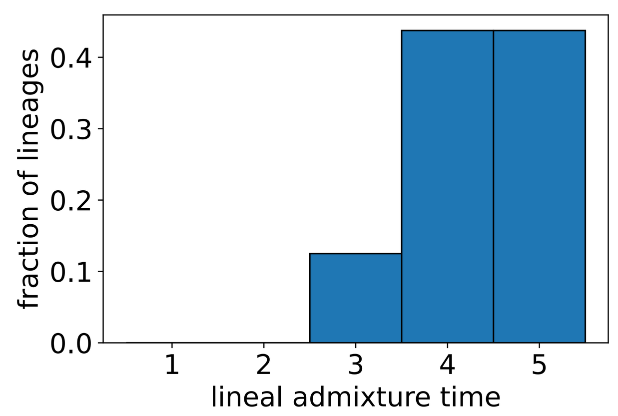
Two scenarios with single pulse of 7 immigrants of red ancestry

Fast admixture
with disassortative mating

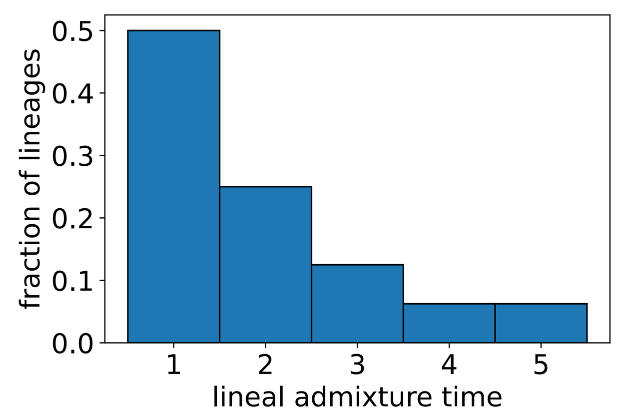
Slow admixture
with assortative mating



Distribution of lineal admixture times across 32 lineages



$\mathrm{P}\{t=3\} = 4/32$
$\mathrm{P}\{t=4\} = 14/32$
$\mathrm{P}\{t=5\} = 14/32$



$\mathrm{P}\{t=1\} = 16/32$
$\mathrm{P}\{t=2\} = 8/32$
$\mathrm{P}\{t=3\} = 4/32$
$\mathrm{P}\{t=4\} = 2/32$
$\mathrm{P}\{t=5\} = 2/32$

Simple estimator

Given

- frequency α_i of alleles from the i-th ancestral group, and
- frequency β of diploid loci with dual ancestry,

estimate **average lineal admixture time** as

$$\frac{1-\phi}{\phi}\left(1 - \Sigma_i x_i^2\right)$$

where

$$x_i = \frac{1 - \sqrt{1 - 4\phi(1-\phi)\alpha_i}}{2(1-\phi)}$$

When there are only two ancestral source populations:

$$\phi = 1 - \frac{\beta}{2\alpha_0(1-\alpha_0)}$$

Learn more at castedo.com/doc/154.

Stochastic process of simple estimator

This simple estimator is precisely the expected lineal admixture time under a stochastic process with the following assumptions:

- discrete time steps
- infinite population
- proportion α_i of immigrants from i-th ancestral group
- fraction ϕ of population is new non-admixed immigrants
- random mating (excluding new immigrants)
- stationary process

The underlying random object of this stochastic process is formally defined as a *gametic lineage*.

For more details, visit castedo.com/doc/153.

Acknowledgements

Special thanks to Hannah Moots and Benjamin Peter for constructive feedback on castedo.com/doc/151 about lineal admixture time and Steven Orzack for mentorship.

Interested in using an advanced admixture time estimator?

Sign up here or castedo.com/poster

Glimpse of estimator derivation

Setup similar to first step analysis:

$$\begin{aligned}\mathrm{E}_{t+1}[M] &= (\mathrm{E}_t[M|M>0]+1)\mathrm{P}_t\{M>0\}^2(1-\phi)\\ &\quad + 2\left(\frac{1}{2}\mathrm{E}_t[M|M>0]+1\right)\mathrm{P}_t\{M>0\}\mathrm{P}_t\{M=0\}(1-\phi)\\ &\quad + \left(\mathrm{P}_t\{M=0\}^2 - \sum_i \mathrm{P}_t\{M=0 \wedge A=e_i\}^2\right)(1-\phi)\end{aligned}$$

Probability of lineal admixture time of zero:

$$\mathrm{P}_{t+1}\{M=0 \wedge A=\mathrm{e}_i\} = \phi\alpha_i + (1-\phi)\mathrm{P}_t\{M=0 \wedge A=\mathrm{e}_i\}^2$$

For more details, visit castedo.com/doc/154.

Castedo Ellerman

castedo.com/about