# Part III: Demographic analysis

- Demographic history includes the history of divergence, fluctuations in effective population size, and gene flow among populations.

- Along with population structure certain demographic histories can result in false positive outliers in selection scans.

- Demographic inference central to studies of biogeography, speciation, conservation, etc.

Best by log-likelihood model. log-likelihood: -8200.48
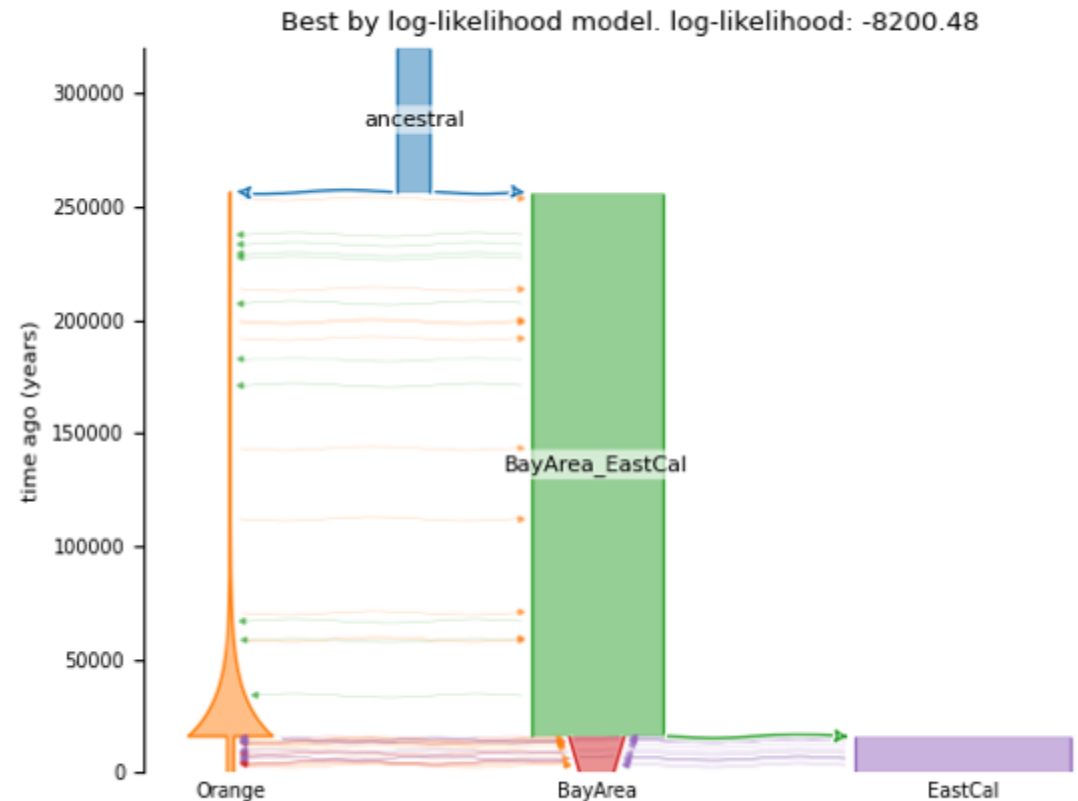
Demographic methods

- Genealogy samplers (e.g. IMa, Lamarc, Migrate) – Kuhner 2009

- G-PhoCS (Gronau et al. 2011)

- Approximate Bayesian computation (ABC)

- Sequentially Markovian Coalescent (e.g. SMC+, PSMC) -- Spence et al. 2018

- Site frequency spectrum based inference (e.g. ðaði, fastsimcoal2)

The site frequency spectrum (1D)
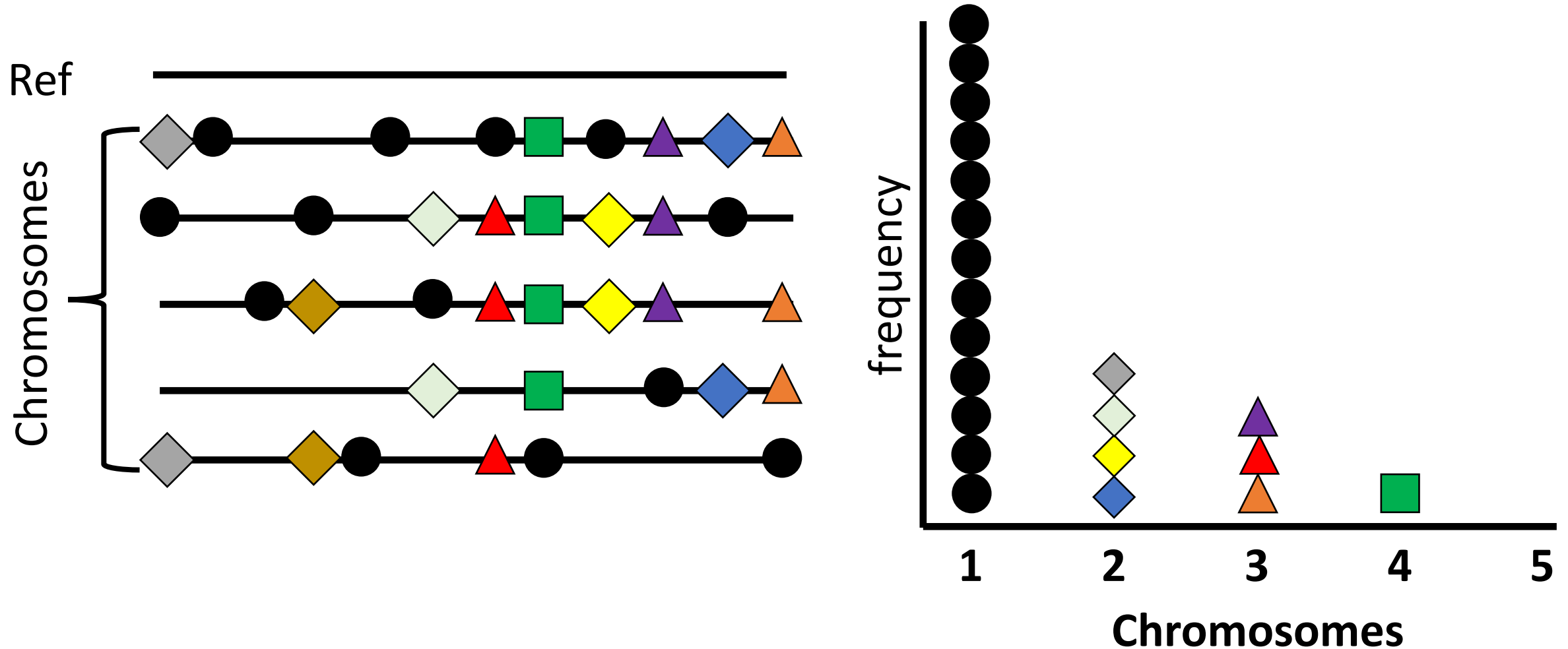
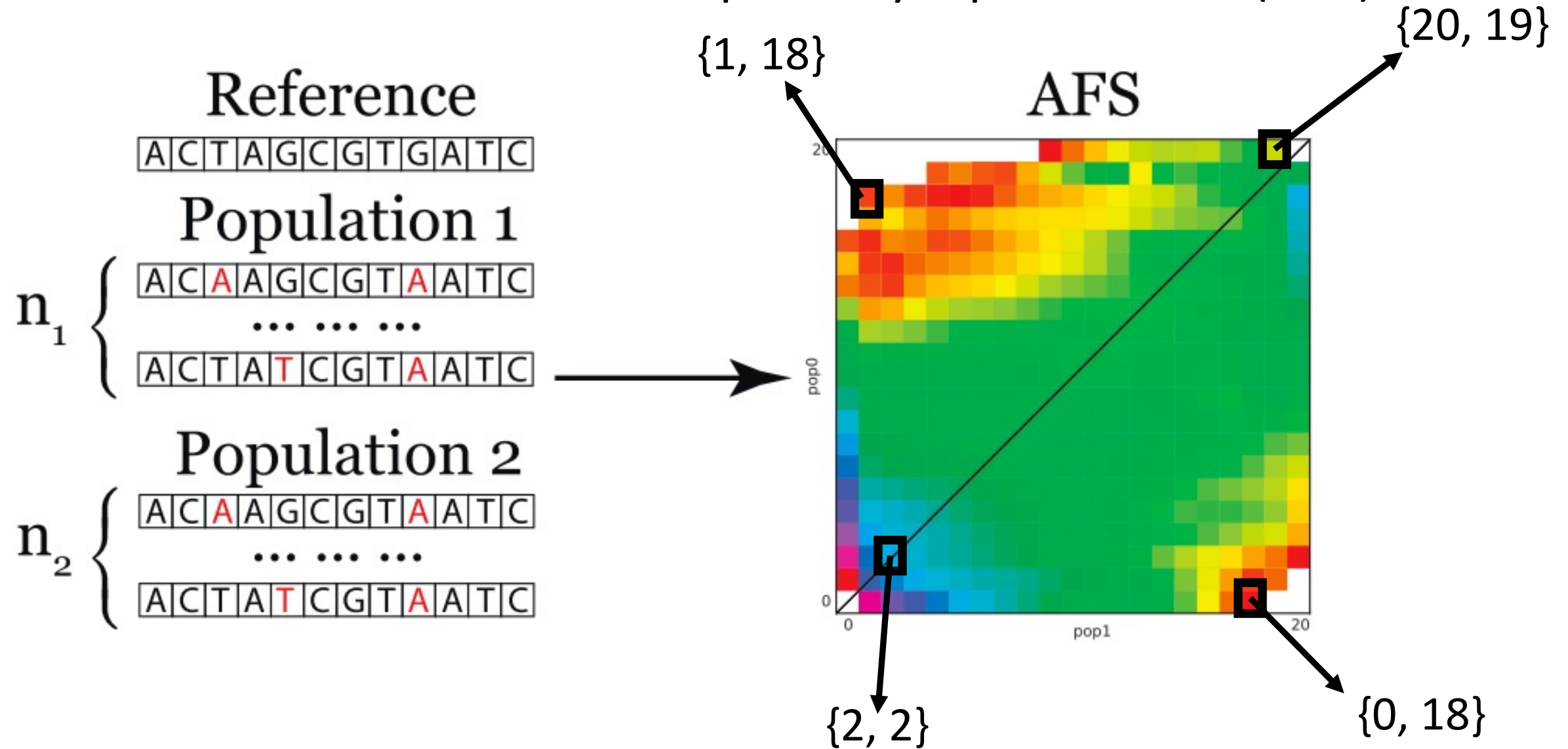# Many pop. gen. statistics are summaries of the SFS
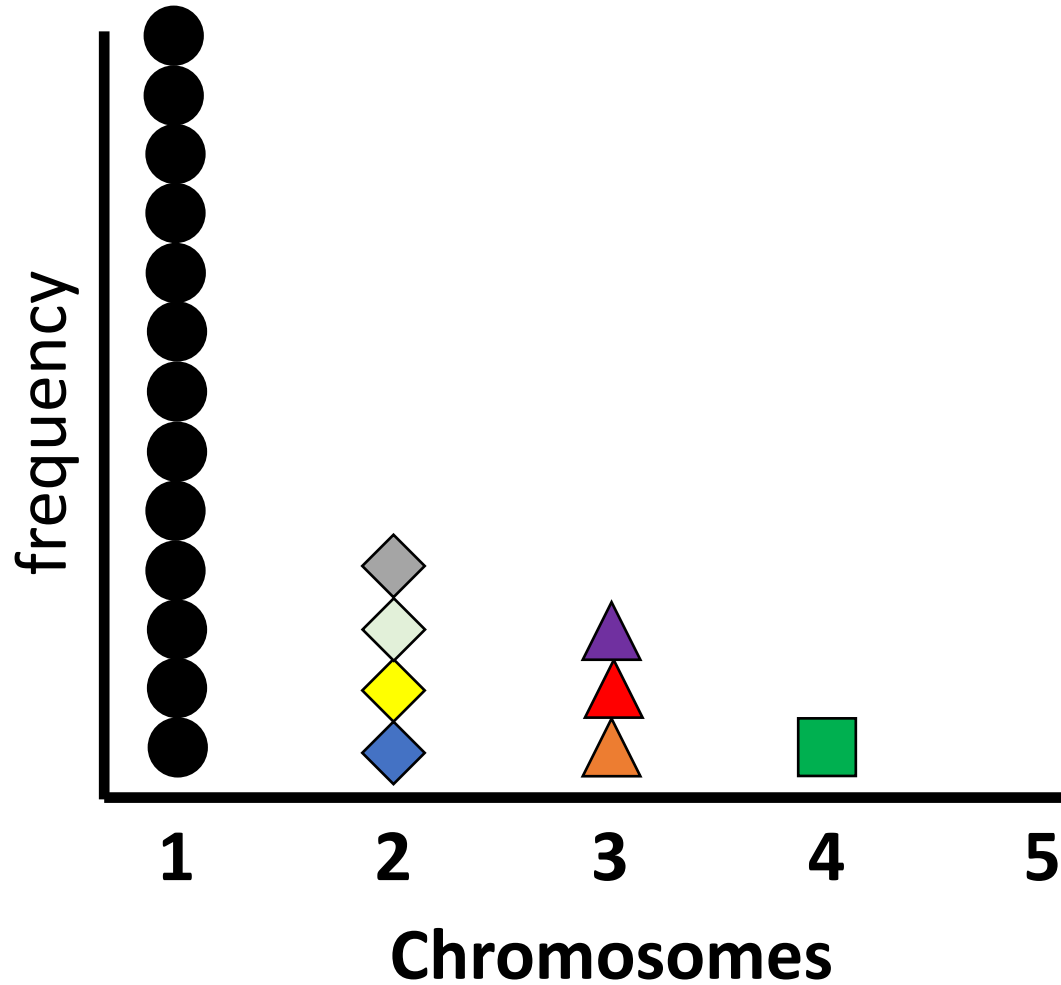
Segregating sites

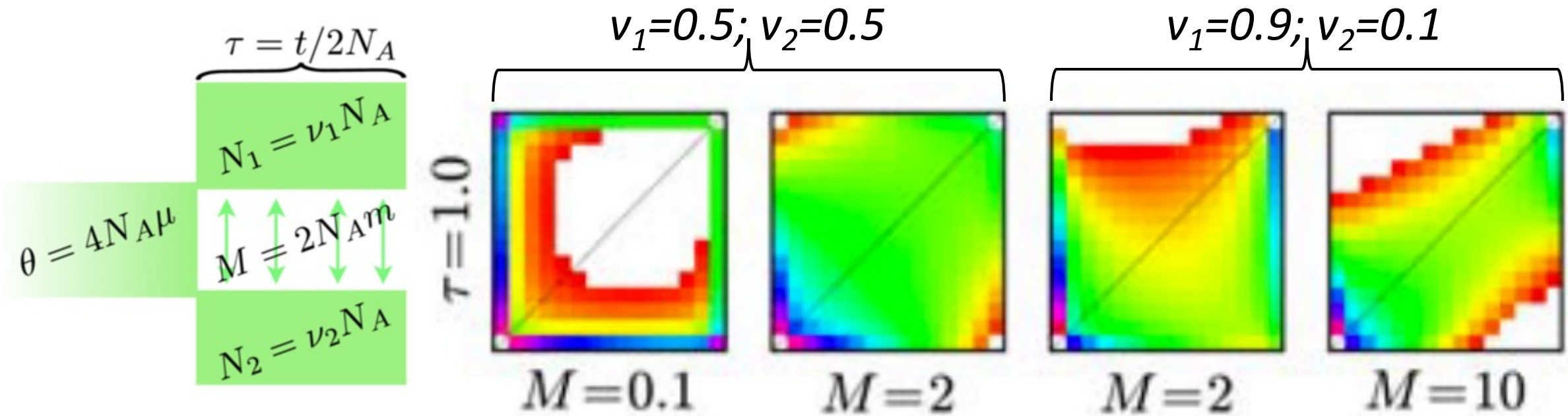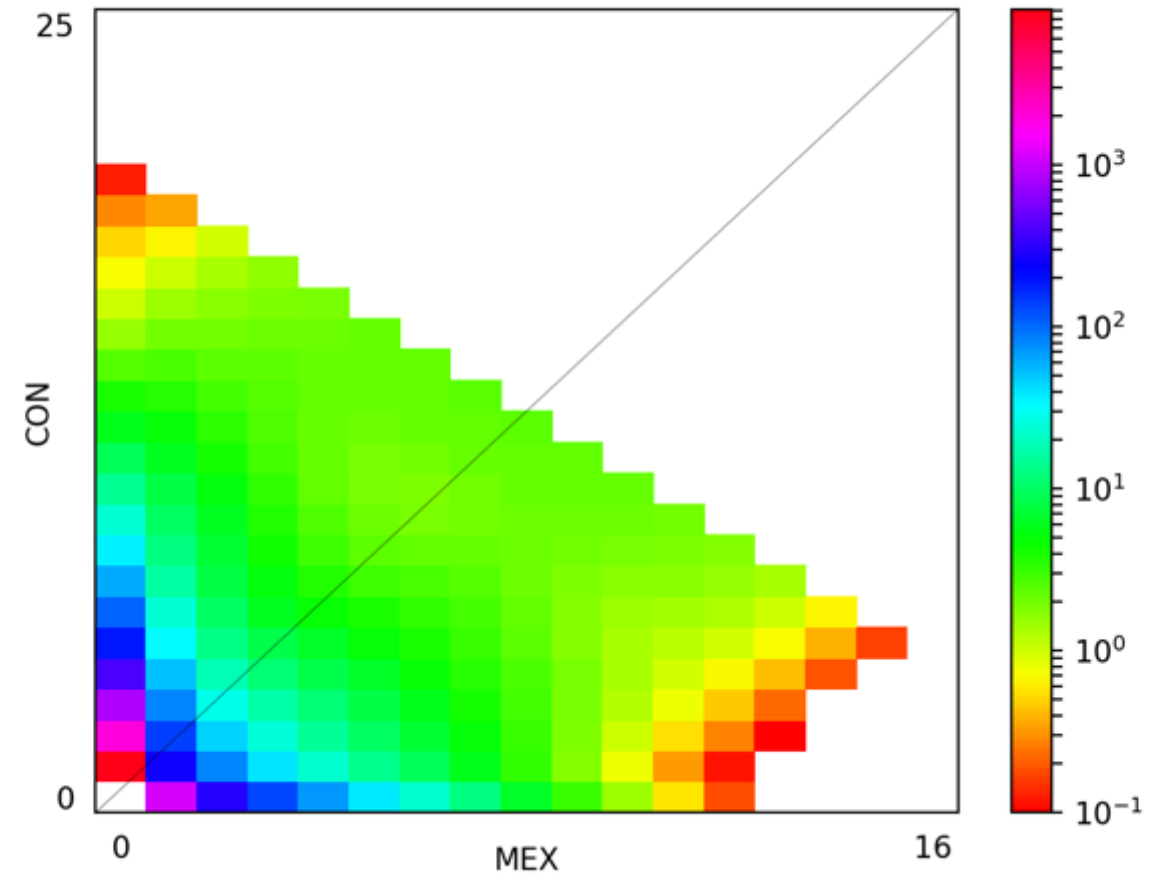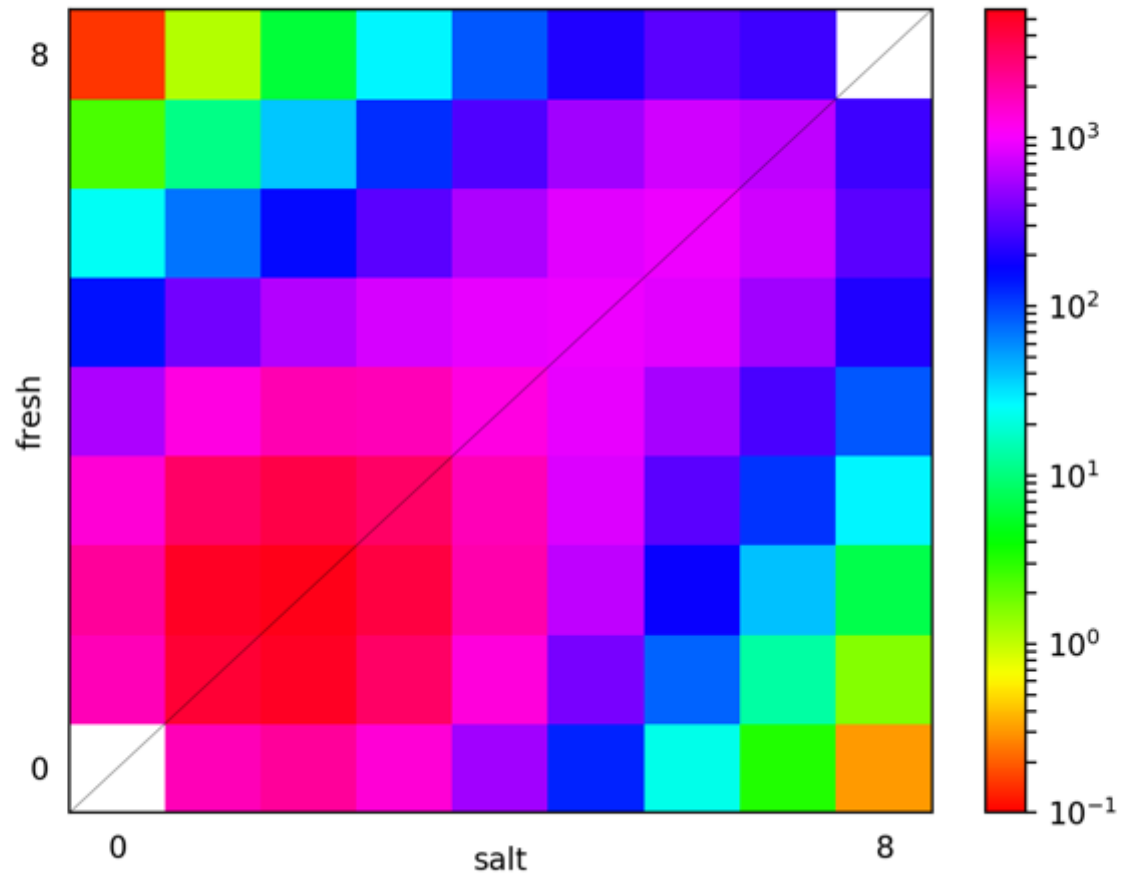Nucleotide diversity

Tajima's D

Joint Site Frequency spectrum (2D)

Reference

A C T A G C G T G A T C

Population 1

A C A A G C G T A A T C

... ...

A C T A T C G T A A T C

n_1

Population 2

A C A A G C G T A A T C

... ...

A C T A T C G T A A T C

n_2

{1, 18}

{20, 19}

AFS

{2, 2}

{0, 18}

pop0

pop1

$$\tau = t/2N_A$$

$$N_1 = \nu_1 N_A$$

$$\theta = 4N_A\mu$$

$$M = 2N_A m$$

$$N_2 = \nu_2 N_A$$

$\nu_1=0.5; \nu_2=0.5$

$\nu_1=0.9; \nu_2=0.1$

$\tau=1.0$

$M=0.1$

$M=2$

$M=2$

$M=10$

Gutenkunst et al. 2009 PLoS Genet.

Polarized vs. folded site frequency spectrum

# Approaches to fitting demographic models to the site frequency spectrum.

- *∂a∂i* – diffusion approximation (Gutenkunst et al. 2009).

- *Moments* (Jouganous et al. 2017).

3-5 populations

- *Fastsimcoal2* – Coalescent based approach (Excoffier et al. 2013).
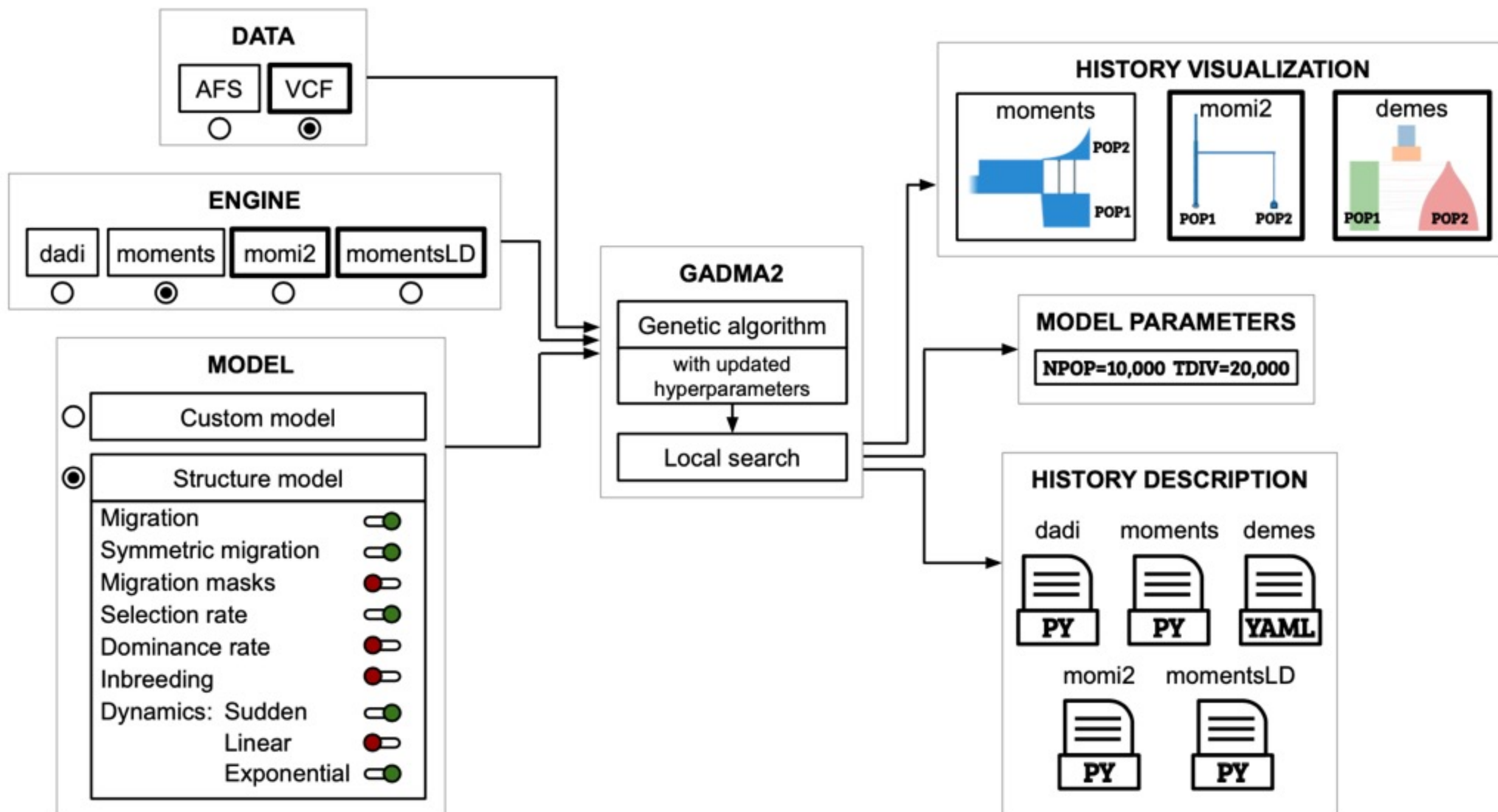
- *Momi* – (Kamm et al. 2017).

Unlimited populations

Pipeline for demographic inference with ∂a∂i and moments
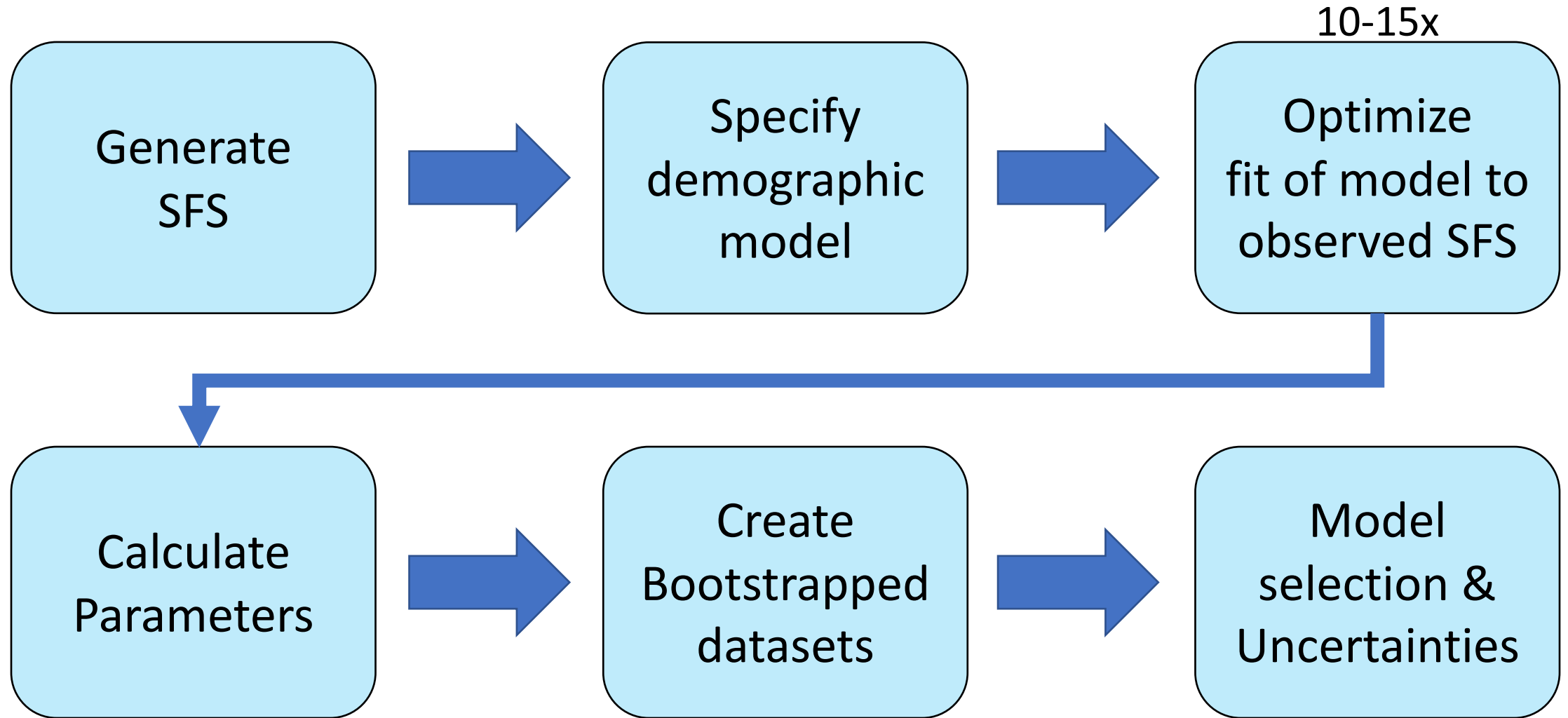
Genetic Algorithm for Demographic Model Analysis
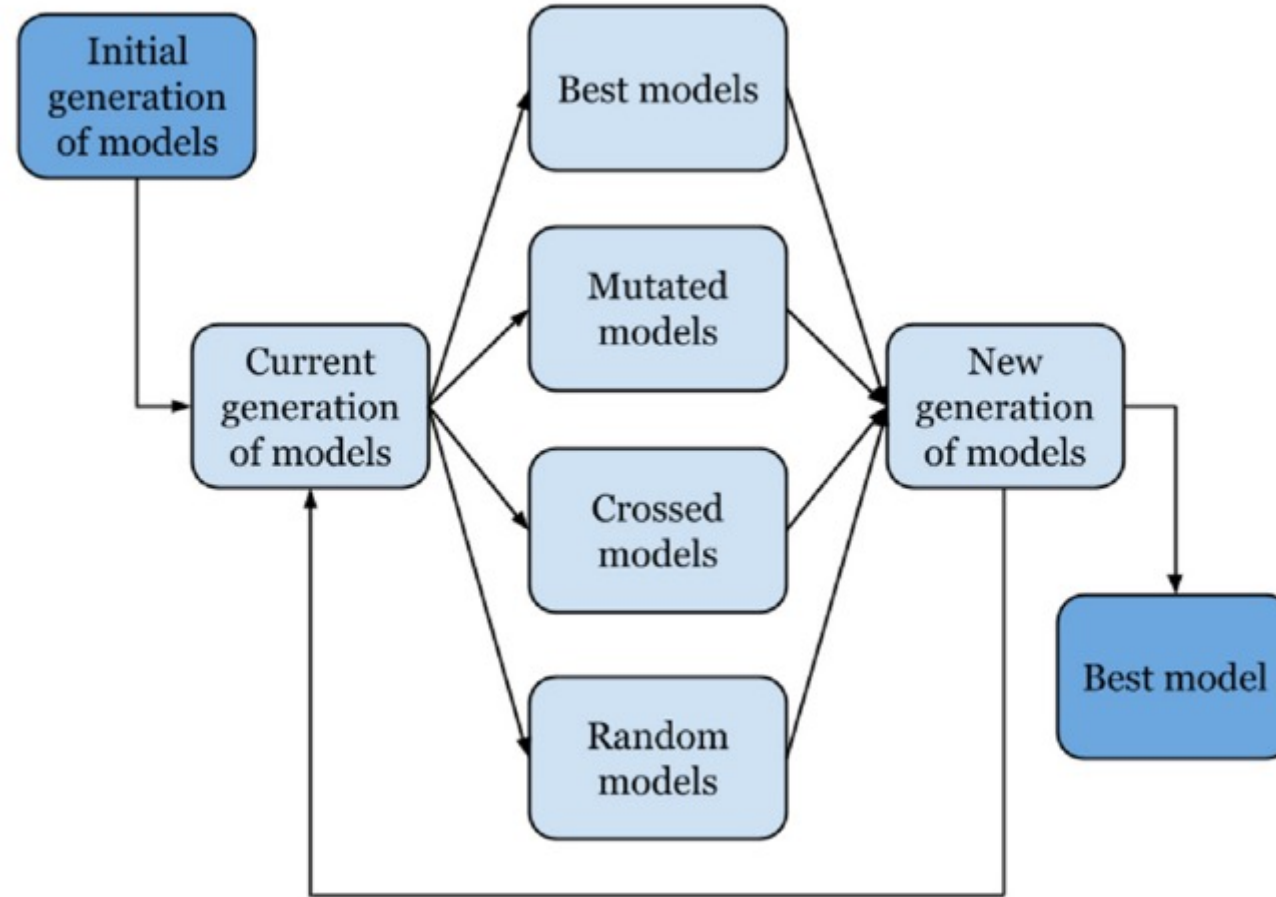
Genetic Algorithm for Demographic Model Analysis

Additional considerations.

- All demographic models are wrong!

- Can be very sensitive to messy data, leads to model misspecification, stacking up against parameter boundaries (e.g., Rosen et al. 2018).

- Bottlenecks can be very challenging to accurately infer. E.g. many datasets underpowered to infer ancient bottlenecks, long versus severe bottlenecks can be conflated (e.g. Terhorst & Song 2015).

- Linked selection causes bias (e.g. smaller population sizes, more recent divergence), non-genic regions preferred for neutral demographic inference (e.g. Schrider et al. 2016).

- Number of individuals: some demographic histories require few individuals (e.g. 5-10) whereas others can require dozens (See Robinson et al. 2014 for some guidelines).

EVEN IN THE FUTURE

NOTHING WORKS

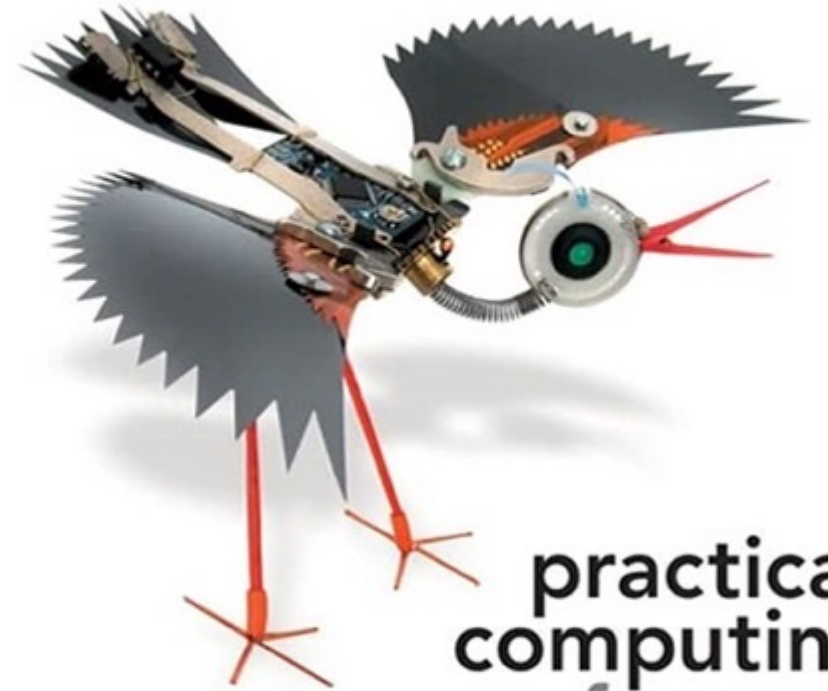made on im

https://methodspopgen.com