

We will explore the files:

- `example_simple_one.py`
- `example_simple_many.py`
- `example_AFR_EUR.py`

together to see some examples how the library `msprime` in `Python` can be used for genomic simulations of different demographic scenarios.

Extend these examples to solve one of the following problems. The material at <https://tskit.dev/msprime/docs/stable/demography.html> or <https://tskit.dev/tutorials/popgen.html> should help you to find the relevant commands.

1. Implement simulations for one of the three models of population size changes for humans (European, Asian, African-American) described in the following figures. (The models with three epochs of constant population size: N_1 , N_2 , and N_3 .) Simulate several datasets of 100kb for a sample of 20 individuals. Compute **diversity**, **Tajima's D**, and the **Site Frequency Spectrum** (allele frequency spectrum) for each simulated datasets, and visualize these values across datasets.

Here is the model:

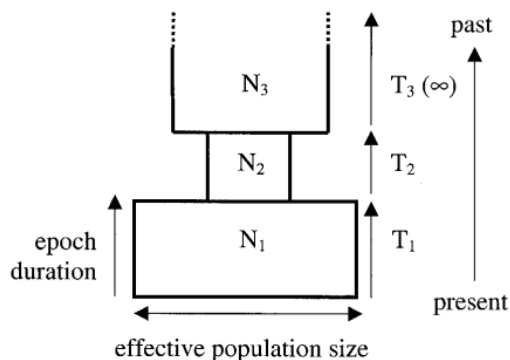


FIGURE 1.—Example of a three-epoch, piecewise constant bottleneck-shaped population history profile. The ancestral effective population size (N_3) is followed by an instant reduction of effective size (N_2). The duration of this epoch is T_2 generations. This is followed by a stepwise increase of effective population size to N_1 , T_1 generations before the present.

And here are the parameters:

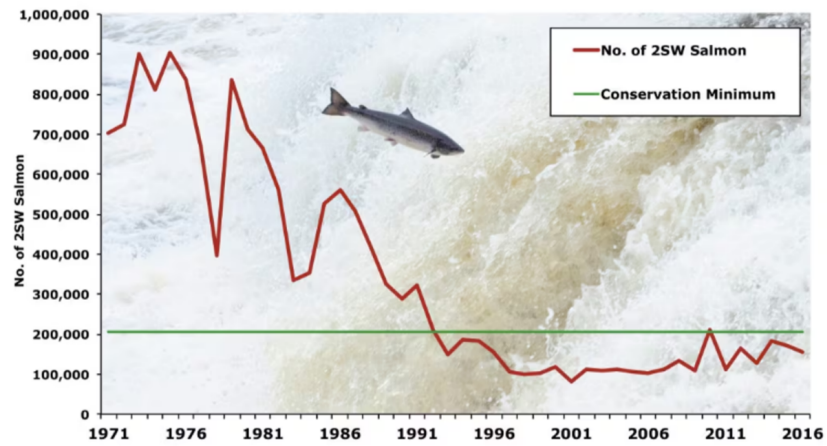
Results of fitting multi-epoch models of allele frequency spectrum to population-specific observed allele frequency data

Model structure	Model parameters	Resulting pairwise θ (units of 10^{-4})	$\ln P(\text{data} \text{model})$	Improvement over lower-epoch model
a. European data				
One epoch	$N_1 = 10,000$	8.00	-55.98	—
Two epoch	$N_2 = 10,000$	8.74	-38.11	$2 \ln \lambda = 35.74$ $P < 10^{-4}$
Three epoch	$N_1 = 140,000$ ($T_1 = 2,000$)	7.88	-23.72	Highly significant
	$N_3 = 10,000$			$2 \ln \lambda = 28.78$
	$N_2 = 2,000$ ($T_2 = 500$)			$P < 10^{-4}$
	$N_1 = 20,000$ ($T_1 = 3,000$)			Highly significant
	b. Asian data			
One epoch	$N_1 = 10,000$	8.00	-74.26	—
Two epoch	$N_2 = 10,000$	8.63	-31.95	$2 \ln \lambda = 84.62$ $P < 10^{-4}$
Three epoch	$N_1 = 50,000$ ($T_1 = 2,000$)	8.24	-26.39	Highly significant
	$N_3 = 10,000$			$2 \ln \lambda = 11.12$
	$N_2 = 3,000$ ($T_2 = 600$)			$P = 0.0039$
	$N_1 = 25,000$ ($T_1 = 3,200$)			Significant
	c. African-American data			
One epoch	$N_1 = 10,000$	8.00	-197.86	—
Two epoch	$N_2 = 10,000$	9.20	-28.69	$2 \ln \lambda = 338.34$ $P < 10^{-4}$
Three epoch	$N_1 = 18,000$ ($T_1 = 7,500$)	10.29	-26.72	Highly significant
	$N_3 = 10,000$			$2 \ln \lambda = 3.94$
	$N_2 = 16,000$ ($T_2 = 15,000$)			$P = 0.1395$
	$N_1 = 26,000$ ($T_1 = 2,400$)			Not significant

2. Implement simulations for the population size history of *Atlantic Salmon* given in the following figure. You can approximate the size history by a piece-wise constant function with 3 pieces (but you can make the approximation more precise later). Assume that one generation is equal to six years. Simulate several datasets of 100kb for a sample of size 20 individuals. Compute **diversity**, **Tajima's D**, and the **Site Frequency Spectrum** (allele frequency spectrum) for each simulated datasets, and visualize these values across datasets.

Here is the model:

Pre-fishery Abundance - North America



The graph shows the decline in the return of wild Atlantic salmon to rivers in North America.
(Submitted/Atlantic Salmon Federation)

<https://www.asf.ca/about-us/>