

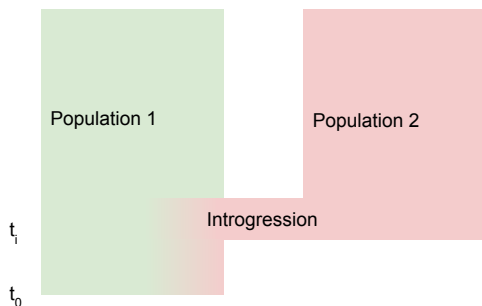
# On the distribution of tract lengths during adaptive introgression

Vladimir Shchur

02 October 2019

# Problem

We are interested in modelling tracts near an introgressed selected allele.



**Figure:** Population 1 (recipient) gets a fraction of population 2 (donor) ancestry through a single pulse introgression.

We assume that there is a selected allele  $A$  (with frequency 1) in the donor population which is not present in recipient population.

## Locus ancestry

- ▶ To trace the ancestry at another locus  $B$ , we need to model the probability that  $B$  is on a chromosome carrying  $A$  at the time of introgression.
- ▶ What is the distribution of length of introgressed tracts overlapping with the locus under selection?

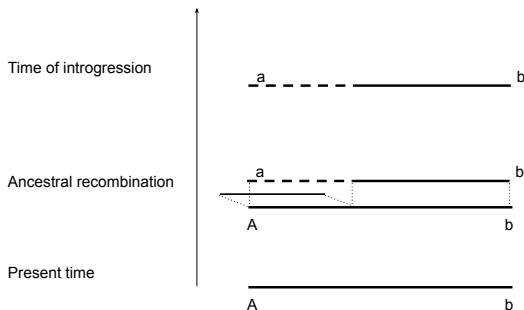
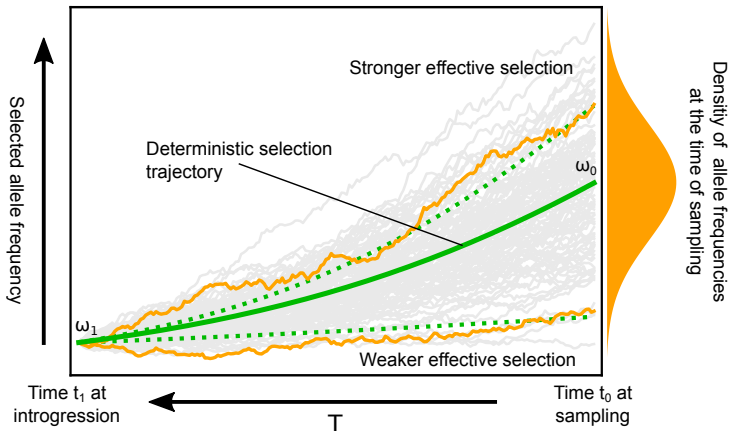


Figure: Locus  $b$  is coming from recipient population.

# Allele frequency: stochastic trajectories



## Allele frequency: logistic trajectory

The expected trajectory of the selected allele is approximated by logistic function

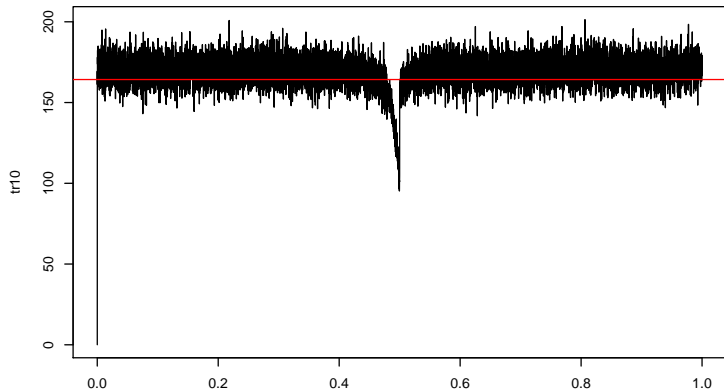
$$\omega(t) = \frac{1}{1 + e^{-st/2}}$$

if

- ▶  $N_e s \gg 1$ ,
- ▶ the allele frequency  $\omega(i) > \varepsilon > 0$  of the selected allele is not close to 0 at the time of introgression  $t_i$ ,
- ▶ the allele frequency  $\omega(0) < 1 - \varepsilon$  of the selected allele is not close to 1 at the present time.

## Transition rates

As one might expect, transition rates (between types of ancestry) are far from being uniform near the selected locus.



**Figure:** Local transition rates from ancestry type 1 to ancestry type 0 (simulated).

# Three loci model

Assume we have a selected site  $a/A$  and two other sites  $b$  and  $c$ .  
There are six different states then:

- ▶  $A-b-c$ ,
- ▶  $A-b^*-c$ ,
- ▶  $A-b^*-c$ ,
- ▶  $a-b^*-c$ ,
- ▶  $a-b-c$ .

This process is described by a system of linear differential equations (backwards in time)

$$\mathbb{P}'(t) = \mathbb{M}(t; r_1, r_2)\mathbb{P}(t).$$

There is no analytical solution for it, but it can be efficiently solved numerically.

## Three loci model

Solving this differential equation for different values of  $r_1$  and some small  $r_2$  leads to an estimate of the transition rate (ancestry type 1 to type 0) given by the following equation

$$\theta_{10}(r) = \lim_{r_2 \rightarrow 0} \frac{1}{r_2} \frac{P(A - b - *, a - * - c)}{P(A - b)}.$$

The expected tract length is

$$\mathbb{E}(\text{tract length}) = 2 \int_0^{\infty} r_1 \theta_{10}(r_1) e^{-\int_0^{r_1} \theta_{10}(r) dr} dr_1.$$



# Accuracy of deterministic approximation

Introgression parameters			Expected tract length			Relative error
Proportion	Selection	Time	Simulations	Deterministic approximation	Integrated	
0.01	0.001	50	0.0417436	0.0401907	0.04027677	3.7%
		100	0.0197123	0.0200985	0.02014723	2.0%
		500	0.00407317	0.00402525	0.004044527	1.1%
		1000	0.00206252	0.00201644	0.0020327	2.2%
0.01	0.01	50	0.0415686	0.0402287	0.04029757	3.3%
		100	0.0202129	0.02014	0.02017217	0.36%
		500	0.00423004	0.00411598	0.004121694	2.8%
		1000	0.00237283	0.00227902	0.002258553	4.0%
0.05	0.001	50	0.0427474	0.0419025	0.04187598	2.0%
		100	0.0210069	0.0209646	0.020949	0.20%
		500	0.00428813	0.00421551	0.004214419	1.7%
		1000	0.00217073	0.00212356	0.00213722	2.2%
0.05	0.01	50	0.0428456	0.0420981	0.04206946	1.7%
		100	0.021438	0.0211766	0.02115895	1.2%
		500	0.00470731	0.00462923	0.004616305	1.7%
		1000	0.00290033	0.00286404	0.002865537	1.3%

Expected tract length estimated from simulations and by deterministic approach.

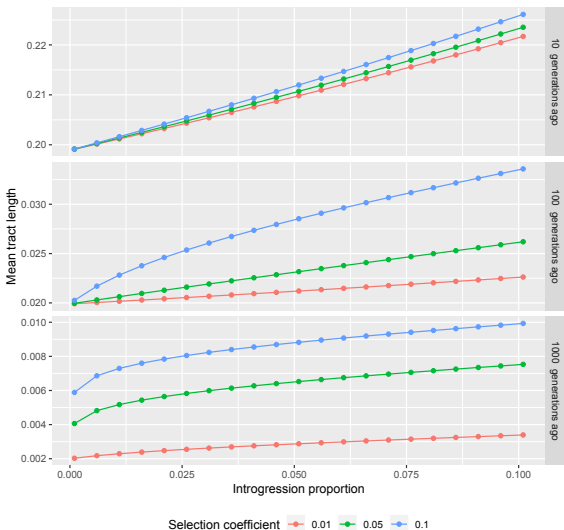
## Neutral case

Even though initially we assume that there should be a selection, in the limiting case of neutral alleles, our approximation still gives consistent results.

Following Liang, Nielsen (Genetics 2014) the expected tract length under introgression without selection can be calculated as

$$\frac{2}{2N_e(1 - \omega_1)(1 - e^{-T/2N_e})}$$

Proportion	Time	Expected tract length (theoretical)	Expected tract length (deterministic approximation)
0.01	10	0.20194	0.200923
	100	0.0201956	0.0200946
	1000	0.00202122	0.00201177
0.05	10	0.210491	0.209387
	100	0.021054	0.0209442
	1000	0.00211024	0.0020999
0.1	10	0.222252	0.222278
	100	0.022233	0.0222778
	1000	0.00223105	0.00227824



**Figure:** Dependence of expected tract length on the proportion of introgression for different times of introgression.

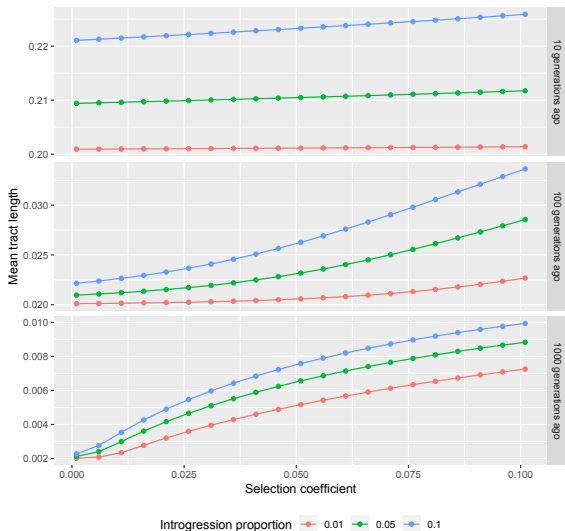
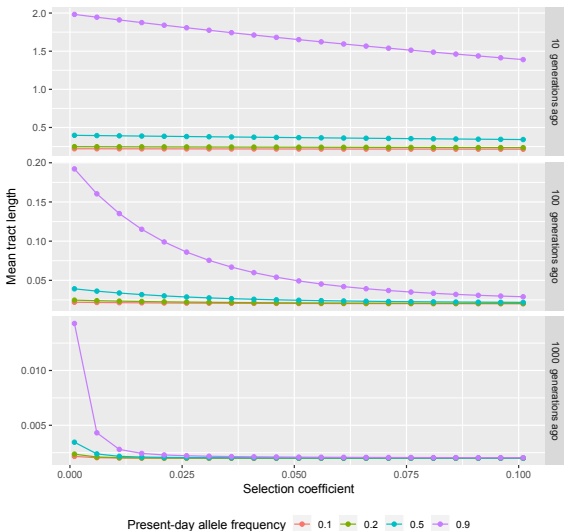


Figure: Dependence of expected tract length on the strength of selection for different times of introgression.



**Figure:** Dependence of expected tract length on the strength of selection conditioned on the present-day allele frequency.

Selection	Time	Expected tract length	Expected tract length (no selection)
0.006	2282	0.00117	0.00093
0.008	1791	0.00146	0.00117
0.01	1467	0.00175	0.00141
0.012	1264	0.00201	0.00163
0.014	1100	0.00230	0.00187
0.016	983	0.00256	0.00209
0.018	883	0.00283	0.00232
0.02	806	0.00309	0.00253

**Table:** The strength of selection considerably affects the dating of the Denisovan introgression into Tibetans.

# Variance

Introgression parameters			Standard deviation of tract length		
Proportion	Selection	Time	Simulated (across populations)	Simulated (within population all replicates)	Deterministic approximation
0.01	0.001	50	0.0293966	0.02740539	0.0285128
		100	0.013484	0.01337553	0.0142598
		500	0.00279459	0.002334261	0.00285811
		1000	0.00145372	0.001131087	0.00143309
0.01	0.01	50	0.0287187	0.02757148	0.0285337
		100	0.0137616	0.01348083	0.0142825
		500	0.0030498	0.002621121	0.00290384
		1000	0.00154791	0.001438002	0.00154791
0.05	0.001	50	0.0304915	0.02958953	0.0297278
		100	0.0149323	0.01464238	0.014875
		500	0.00303497	0.002756722	0.00299345
		1000	0.00155372	0.001336331	0.00150908
0.05	0.01	50	0.0306196	0.02973097	0.0298345
		100	0.0149798	0.01483033	0.0149894
		500	0.00324606	0.003119859	0.00319587
		1000	0.0018881	0.001784578	0.00183546

# Variance

Introgression parameters			Standard deviation of tract length		
Proportion	Selection	Time	Simulated (across populations)	Simulated (within population all replicates)	Deterministic approximation
0.0006	0.01	1500	0.001173	0.000962	0.001125
		1750	0.000998	0.000889	0.001014
		2000	0.000954	0.000811	0.000930
		2250	0.000884	0.000751	0.000864
0.0006	0.02	1500	0.001423	0.001241	0.001350
		1750	0.001312	0.001132	0.001230
		2000	0.001221	0.001050	0.001136
		2250	0.001179	0.000974	0.001061
0.025	0.001	1500	0.001052	0.000806	0.001015
		1750	0.000840	0.000677	0.000878
		2000	0.000835	0.000589	0.000776
		2250	0.000691	0.000535	0.000696
		2500	0.000608	0.000477	0.000633
0.025	0.005	1500	0.001136	0.000991	0.001082
		1750	0.000970	0.000875	0.000956
		2000	0.000887	0.000784	0.000863
		2250	0.000836	0.000726	0.000793
		2500	0.000779	0.000672	0.000737



# Collaborators

- ▶ Rasmus Nielsen (UC Berkeley)
- ▶ Russ Corbett-Detig (UC Santa Cruz)
- ▶ Jesper Svedberg (UC Santa Cruz)
- ▶ Paloma Medina (UC Santa Cruz)