Mitochondrial DNA and the History of Population Size

Alan R. Rogers

September 2, 2021

Why mitochondrial DNA is simple

- inherited from mother only
- no recombination
- evolves fast
- mean pairwise difference: average number of nucleotide site differences between mitochondria of pairs of individuals
- we will look at about 300 base pairs of mitochondrial DNA

Mitochondrial nucleotide differences (D-loop)

Mean pairwise diff: between\within\net

Major Human Populations

	European ($N = 20$)	Asian ($N = 71$)	African ($N = 10$)
Eur	**0.0094**	0.0012	0.0028
As	0.0128	**0.0137**	0.0015
Af	0.0194	0.0203	**0.0238**

Chimpanzee Subspecies

	Pts ($N = 40$)	Ptt ($N = 18$)	Ptv ($N = 8$)
Pts	**0.029**	0.030	0.137
Ptt	0.072	**0.055**	0.087
Ptv	0.192	0.155	**0.081**

Why are genetic differences between Europeans and between Asians smaller than those between Africans?

Why are differences within human populations smaller than those within chimpanzee subspecies?

What determines the amount of genetic variation within a population?

Let's examine the effect of population size.

Genetic drift: random change in allele frequency

Some people die early, just by chance.

Some have more children than others, just by chance.

Each time you produce a gamete, there is a 50:50 chance that it will carry at some site the nucleotide you got from Mom or the one from Dad.

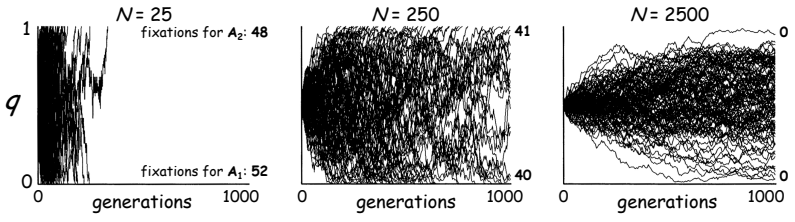
These things cause random changes in allele frequencies—genetic drift.

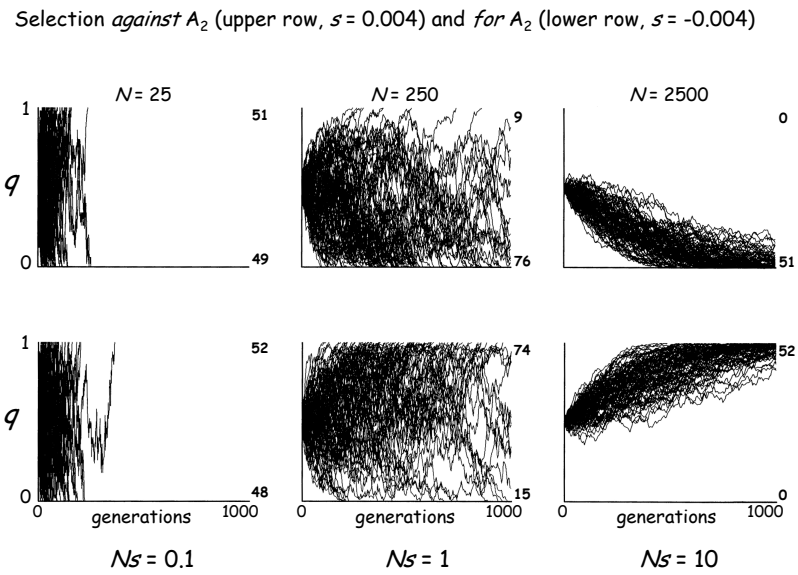
In large populations, random effects tend to average out, so the effect of drift is small. Its effect is large in small populations.

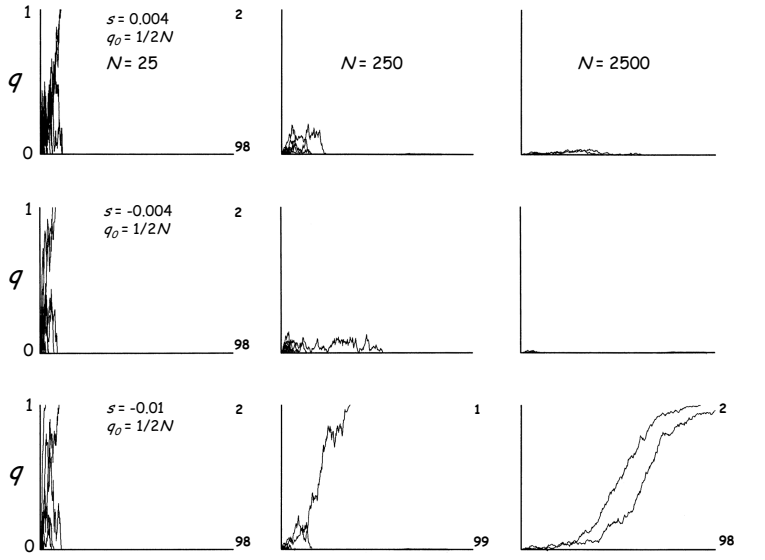
Movies of drift versus selection (N, s, and the fates of mutations)

Each panel shows 100 allele-frequency histories for a given N, s, and initial frequency q_0.
Each history runs for 1000 generations or until fixation occurs.
Within each row of three panels, the population sizes are 25, 250, and 2500.
$W(A_1A_1) = 1$, $W(A_1A_2) = 1-\frac{1}{2}s$, and $W(A_2A_2) = 1-s$.
The first set of cases (below) is a neutral "control" for the experiments in subsequent slides.

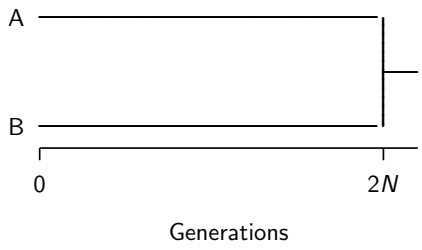
$s = 0$, $q_0 = \frac{1}{2}$ (neutral case, pure drift)



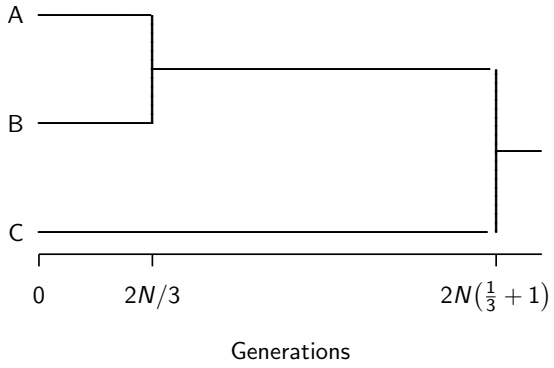




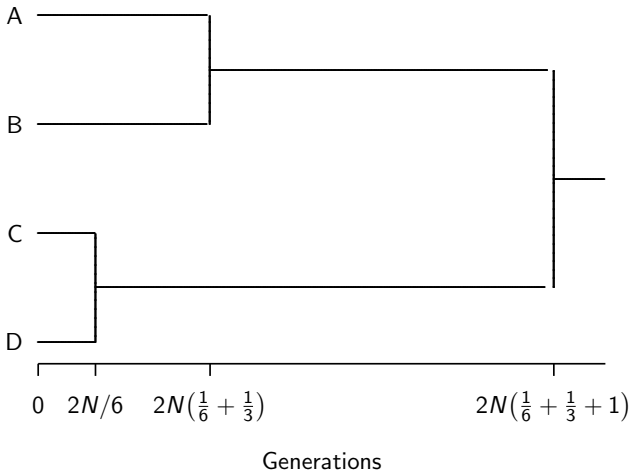
Genealogy of 2 genes



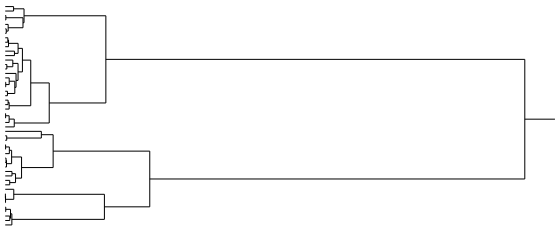
Genealogy of 3 genes



Genealogy of 4 genes



Simulated gene genealogy of a sample of size 50 from a population of constant size



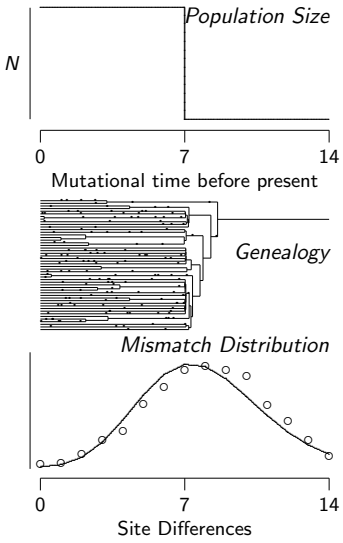
- Terminal branches are short.
- Basal branch is long.
- Additional samples don't tell us much about the ancient past.

Principles

The expected length of a coalescent interval is long

- in large populations
- if there are only a few lineages

What if the population changes in size?



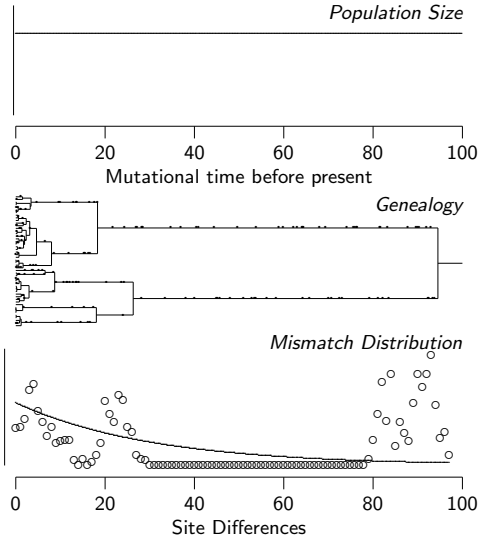
Effect of a population explosion

Middle: comb-shaped (or star-shaped) genealogy of 50 individuals; dots are mutations.

1 mutational diff per time unit

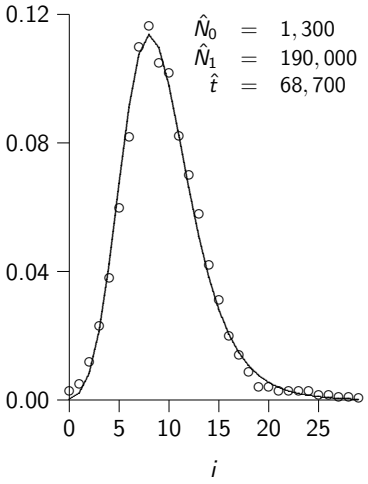
Bottom: ○ = simulated data, line = theory.

Wave peaks at population expansion.



- Simulation of equilibrium population
- No history of growth
- Deep genealogy
- Ragged mismatch distribution

Mitochondrial Mismatch Distribution



Open circles: data of Cann, Stoneking, & Wilson (1987).

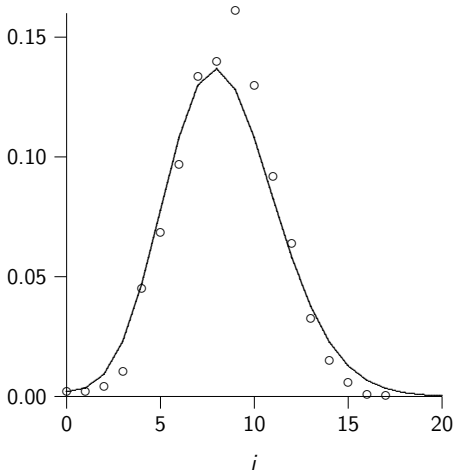
Solid line: fit to expansion model (Rogers & Harpending 1992).

Implies expansion ∼70 kyr ago (or so we thought then).

Revised molecular clock: ∼45 kyr ago.

Coincident with expansion of modern humans into Eurasia.

Mismatch Distribution: 77 Asian Subjects



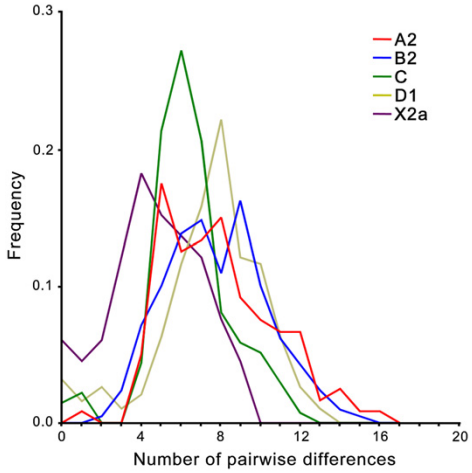
Open circles: Asian data

Solid line: fit to expansion model (Rogers & Harpending 1992)

Implies expansion ∼45 kyr ago

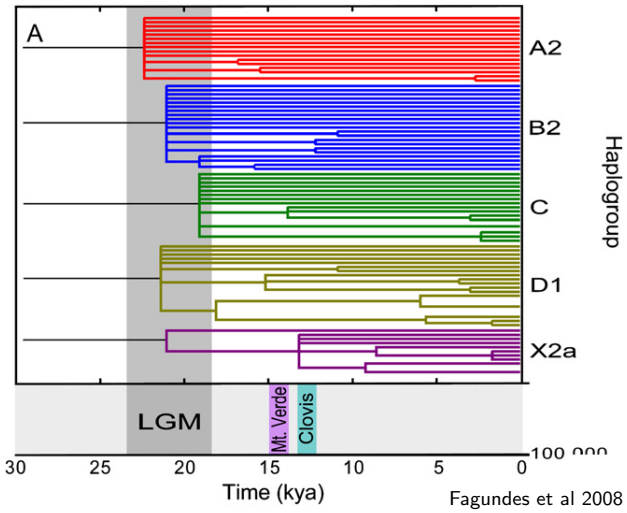
Coincident with expansion of modern humans into Eurasia.

Mismatch Distributions of Amerindian mtDNA Haplogroups



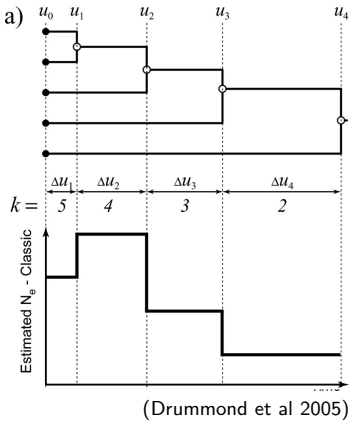
Fagundes et al 2008

Genealogies of Amerindian mtDNA Haplogroups



Fagundes et al 2008

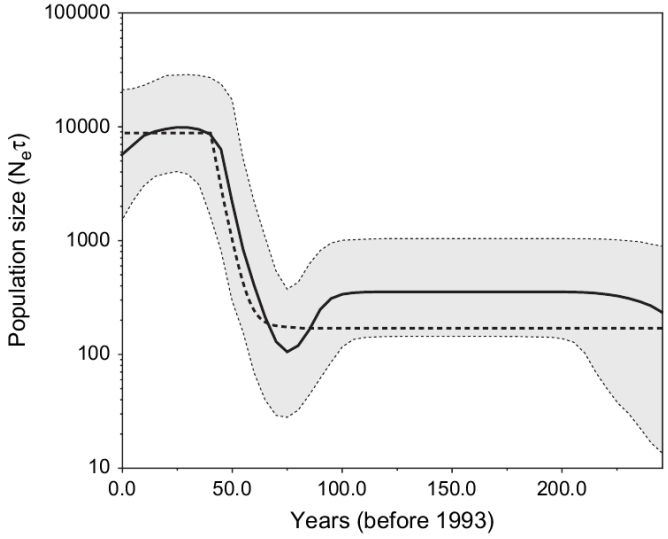
Skyline Plot



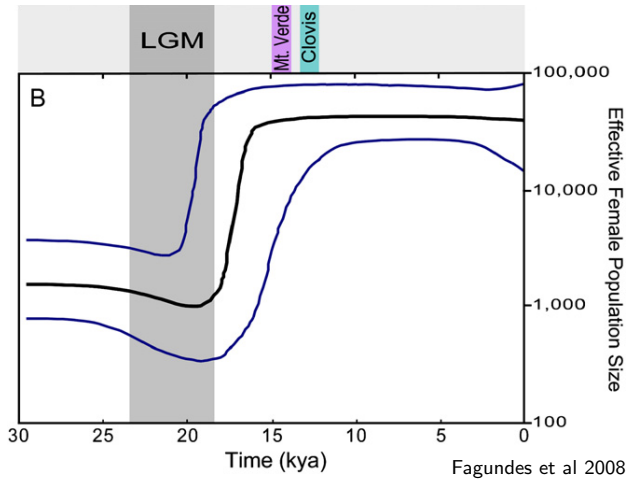
(Drummond et al 2005)

- Use mutations to estimate length of each interval.
- Long intervals imply large population size.
- Won't work with nuclear DNA: too few mutations per tree

Hepatitus C virus in Egypt (Drummond et al 2005)

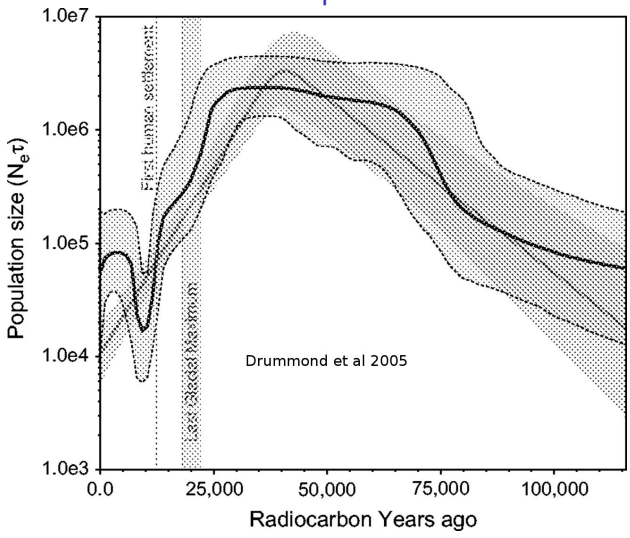


Estimated Size of Amerindian Population



Fagundes et al 2008

Estimated Size of Bison Population



The Selection Hypothesis

Suppose that a favorable mitochondrial mutation

- arose 60,000 years ago, and
- increased in frequency until everyone had it.

This is called a "selective sweep."

Effect on mitochondrial variation is indistinguishable from that of population growth.

Problems with mitochondrial clock

- No recombination: deleterious mutations are hard to get rid of.
- Hang around in mtDNA longer than in nDNA.
- At short time scales, we overestimate the count of neutral differences.
- Overestimate dates within past 200,000 years.

Summary

- History of population size affects depth of gene trees, genetic variation, and length of MRCA segments.
- We can use these facts to infer the history of population size.
- Mitochondrial DNA suggest that human populations were small during the last ice age.
- With mitochondrial DNA, it is hard to distinguish the effect of population growth from that of selection.