The History of Population Size from Whole Genomes

Alan R. Rogers

October 1, 2018

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

- Short terminal branches; long basal ones.
- Large samples tell us about recent past.
- Not necessary for ancient past.

Skyline Plot

- 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

Nuclear genome

- Huge amounts of data.
- Recombination makes previous methods unusable.

But we can still use the site frequency spectrum

Population size

Mutational time before present

Site frequency spectrum

Frequency of minor allele

Wave peaks at population expansion.

Effect of a population explosion

Middle: genealogy of 50 individuals; dots are mutations.
1 mutational diff per time unit
Bottom: ○ = simulated data, line = theory.

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

Nuclear genome

- Huge amounts of data.
- Recombination makes previous methods unusable.

But we can still use the site frequency spectrum
\( \partial a \partial i \): inferences from the site frequency spectrum

\[ N_A \quad N_{AF} \]
\[ T_A \quad T_B \]
\[ m_{AF-B} \quad m_{AF-EU} \]
\[ N_{EU0} \quad N_{EU} \quad m_{EU-AS} \]
\[ N_{AS0} \quad r_{AS} \]
\[ T_{EU-AS} \quad (also \quad m_{AF-AS}) \]

(Gutenkunst et al 2009)

Stairway Plot uses spectrum

1000-Genomes data; (Liu & Fu 2015)

Also: recombination is our friend

Two (hypothetical) loci in a single diploid genome


Crossovers shuffle DNA

Each chromosome has many gene genealogies, which vary in length.

Circles: nucleotide sites that differ (are heterozygous) in a single diploid sample.

Heterozygous sites are denser where gene tree is deep.

Population size \( \rightarrow \) length of MRCA segments and genetic variation within segments.

PSMC uses this pattern to estimate population history

PSMC is accurate from 30 ky to 3 my ago.

(Li and Durbin 2011)
PSMC estimates from autosomes

↑ 2 mya (origin of Homo);
↑ 200 kya (origin of modern humans); ↑ 20 kya (beginning of Holocene).

Eurasian/African split 150 kya.
African bottleneck short and shallow.

PSMC estimates from X chromosomes

MSMC: using multiple genomes

PSMC with Neanderthal as well as Denisova

Population tree
Gene genealogies and nucleotide site patterns

- Gene genealogy within population tree.
- Mutation on red branch → site pattern \( yn \).
- Blue branch → \( ynd \).
- \( yn \): 0, ancestral; 1, derived.
- \( ynd \): 0, ancestral; 1, derived.

The mystery of the 4000-year-old Denisovan

We argued in 2017 for an early separation of Neanderthals and Denisivans and a bottleneck among their ancestors. Mafessoni and Prüfer showed that results are different if one includes singleton site patterns. However, the with-singleton analysis also implied an implausible 4 kya date for the Denisovan fossil. How can this be explained?

Observed Site Pattern Frequencies

- \( xy \): 0 0 0 0
- \( xn \): 1 1 1 1
- \( xd \): 0 0 0 0
- \( yn \): 1 1 1 1
- \( yd \): 0 0 0 0
- \( nd \): 0 0 0 0
- \( yx \): 1 1 1 1
- \( yd \): 0 0 0 0
- \( nd \): 0 0 0 0
- \( yd \): 0 0 0 0

Site Pattern Frequency (fraction of nucleotide sites exhibiting each pattern)

- \( X \), Africa; \( Y \), Europe; \( N \), Neanderthal; \( D \), Denisovan.
- \( xy \) is common because \( X \) and \( Y \) share ancestry.
- Ditto \( nd \).
- Goal: infer history from these data

Clues: an excess of site patterns \( d \) and \( xyn \)

- \( dx \): 0 0 0 0
- \( nx \): 1 1 1 1
- \( xd \): 0 0 0 0
- \( nd \): 0 0 0 0
- \( xyn \): 1 1 1 1
- \( yxd \): 0 0 0 0
- \( xnd \): 0 0 0 0
- \( ynd \): 0 0 0 0

Site Pattern Frequency

- Suggests hyperarchaic admixture into Denisovans (Prüfer et al. 2014) or early modern admixture into Neanderthals (Kuhlwm et al. 2016).

Two ways to inflate \( d \) and \( xyn \).

- Red branch mutations generate site pattern \( d \); Blue generates \( xyn \).
- \( d \): 0 0 0 0
- \( xyn \): 1 1 1 1

Admixture Fraction

- \( m_{XY} \)
- \( m_{h} \)
- \( m_{n} \)

Haplod Population Size (2N)

- \( 2N_{XY} \)
- \( 2N_{h} \)
- \( 2N_{n} \)

10^5 Years
### Implications of previous slide

- \( m_{H}, T_{XY} \): Substantial \( H \rightarrow D \) admixture and a hyperarchaic separation time of \( \sim 1.6 \text{ mya} \).
- \( m_{XY} \): \( XY \rightarrow N \) admixture.
- \( T_{ND} \): Early separation of Neanderthals and Denisovans: \( \sim 747 \text{ kya} \).
- \( 2N_{ND} \): Narrow bottleneck in Neanderthal-Denisovan ancestors.
- \( 2N_{AV}, 2N_{N} \): Effective Neanderthal population was large early \( (2N_{AV} \approx 21 \text{ k}) \) but small later \( (2N_{N} \approx 5 \text{ k}) \).
- \( T_{V}, T_{A}, T_{D} \): Vindija, Altai, and Denisovan fossil ages are \( \sim 70 \text{ ky}, \sim 150 \text{ ky}, \text{ and } \sim 100 \text{ ky} \).

### Summary (part 1)

- 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.
- Human population has varied in size over past 3 my.
- Bottleneck during last ice age, ending 20 kya.
- African bottleneck was shorter and shallower.
- Eurasian/African split 150 kya.
- European/Asian split 20 kya.

### Summary (part 2)

#### Current consensus
- Neanderthals and Denisovans separated \( \sim 450 \text{ kya} \),
- then declined to tiny population sizes (\(<1000\) individuals).

#### Our view
- Archaics separated from moderns 750 kya,
- then endured a bottleneck of \( \sim 5 \text{ ky} \).
- Neanderthals & Denisovan separated shortly thereafter.
- Neanderthal population was large early & small later.