How old is the ancestor of two alleles?

Nick J. Patterson

Broad Institute
July 9 2004

Intend to revisit some old results of Kimura
Take point of view of coalescent.

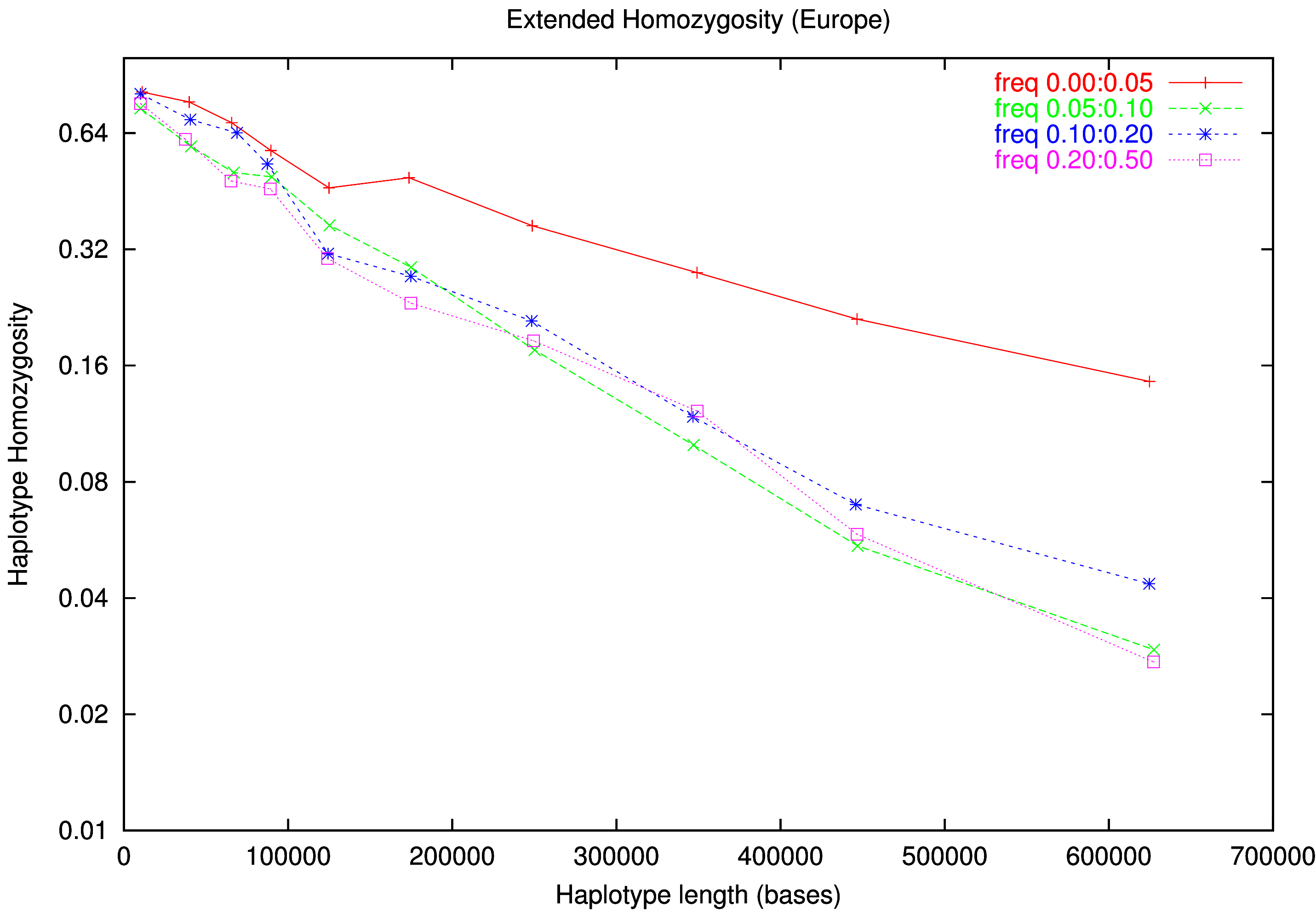
- Motivation: Some actual data!

- The diffusion equation

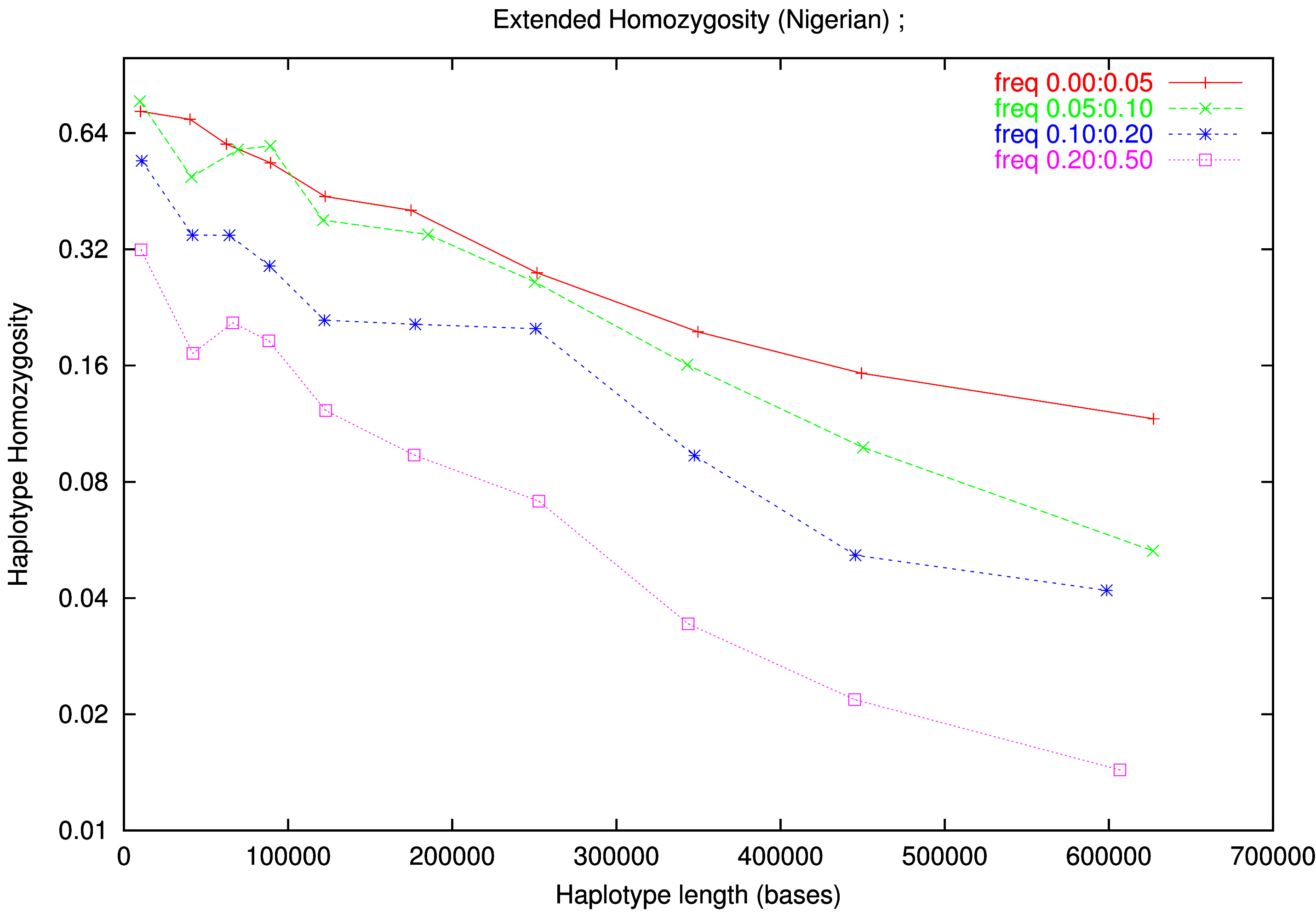
- An analytic result

- A Martingale argument

- Future work?



Extended Homozygosity (Europe)
freq 0.00:0.05
freq 0.05:0.10
freq 0.10:0.20
freq 0.20:0.50
Haplotype Homozygosity
0.64
0.32
0.16
0.08
0.04
0.02
0.01
0
100000
200000
300000
400000
500000
600000
700000
Haplotype length (bases)



Extended Homozygosity (Nigerian) ;
freq 0.00:0.05
freq 0.05:0.10
freq 0.10:0.20
freq 0.20:0.50
Haplotype Homozygosity
0.64
0.32
0.16
0.08
0.04
0.02
0.01
0
100000
200000
300000
400000
500000
600000
700000
Haplotype length (bases)

Panmictic constant size population
Study coalescent and allele frequency jointly
Unit time $2N$ generations.

Let us look at 2 haplotypes carrying an allele
If x is allele frequency at $-t$
define $C(x, y, t)$ to be
joint pdf that frequency is y at time 0
and no coalescence has occurred

C is main object of study.

Backwards diffusion equation: *(Griffiths)*

$$\frac{\partial C}{\partial t} = \frac{x(1-x)}{2}\frac{\partial^2 C}{\partial x^2} - \frac{C}{x} \tag{1}$$

There is also a forwards equation.

This is the Wright-Fisher diffusion with *killing*.

(k copies of an allele)
$C_2 = C$
Theorem

$$C_k(x,y;t) = y^{k-1}x^k(1-x). \sum_{i=0}^{\infty} \frac{J_i^{1,2k-1}(x)J_i^{1,2k-1}(y)}{N_i^{1,2k-1}} e^{-\lambda_k(i)t}$$

where

$$\lambda_k(i) = \frac{(i+k)(i+k+1)}{2}$$

The polynomials $J_i^{1,2k-1}$ are *Jacobi polynomials*
and $N_i^{1,2k-1}$ are explicit constants

Generalizes a result of Kimura for $k=1$.

The series behaves badly for small t (Gibbs Phenomenon)
Taylor expansion is available.

Theorem

Let an allele have frequency f.
Consider k copies of the allele.

Expected time to first coalescent event is $\frac{f}{\binom{k}{2}}$

Expected age of each coalescent event same as in ordinary coalescent (with fixed size population f)

Result not as obvious as it may seem!

(Idea of proof)
From diff. eqn.
show that if X is frequency at coalescence
and τ is coalescence time then:

$$E(X) = \binom{k}{2} E(\tau)$$

But *(Watterson)*
time-reversed process is Martingale
and τ is stopping time.
Thus $E(X) = f$ and result follows

Future work

- Run exact coalescent simulator keeping track of population frequency.

- Generalize to case of selection

Acknowledgments

David Reich (constant encouragement)
Noam Elkies (help with a proof)
Bob Griffiths (excellent advice)