

# Gene trees and phylogeography

[illegible]

Scott V. Edwards  
Department of Organismic  
and Evolutionary Biology  
Harvard University

# Gene trees and phylogeography

- What is phylogeography?
- How can we measure genetic variation within a species?
- What can gene trees tell us about population history?
- How can we infer natural selection from population genetic data?

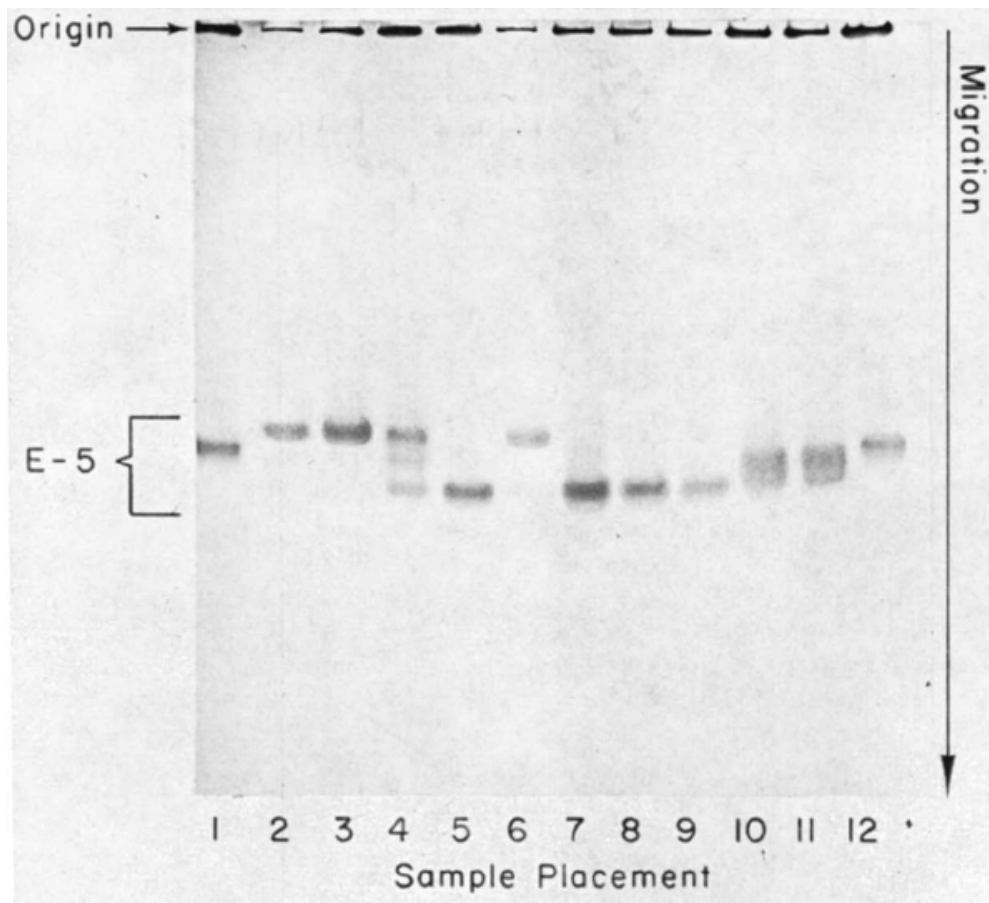
A MOLECULAR APPROACH TO THE STUDY OF GENIC  
HETEROZYGOSITY IN NATURAL POPULATIONS.

I. THE NUMBER OF ALLELES AT DIFFERENT  
LOCI IN *DROSOPHILA PSEUDOOBSCURA*<sup>1</sup>

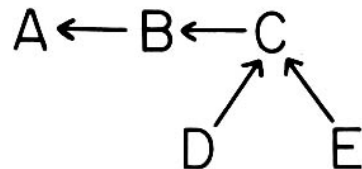
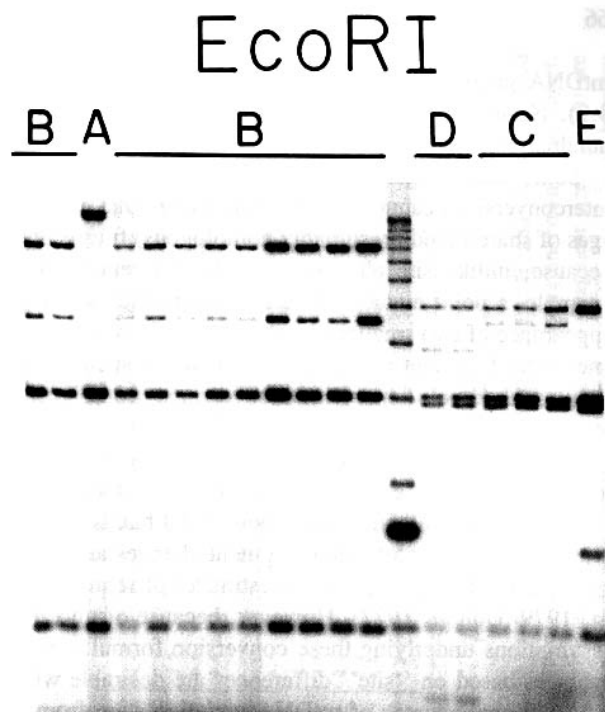
J. L. HUBBY AND R. C. LEWONTIN

*Department of Zoology, University of Chicago, Chicago, Illinois*

Received March 30, 1966

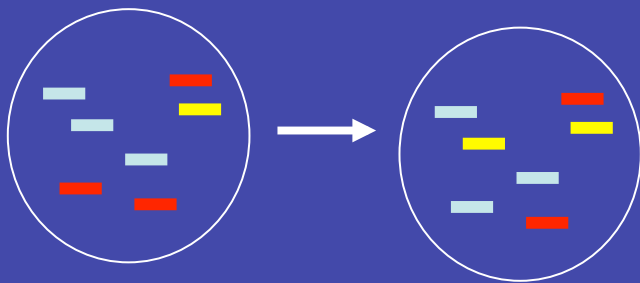


# Restriction enzyme analysis



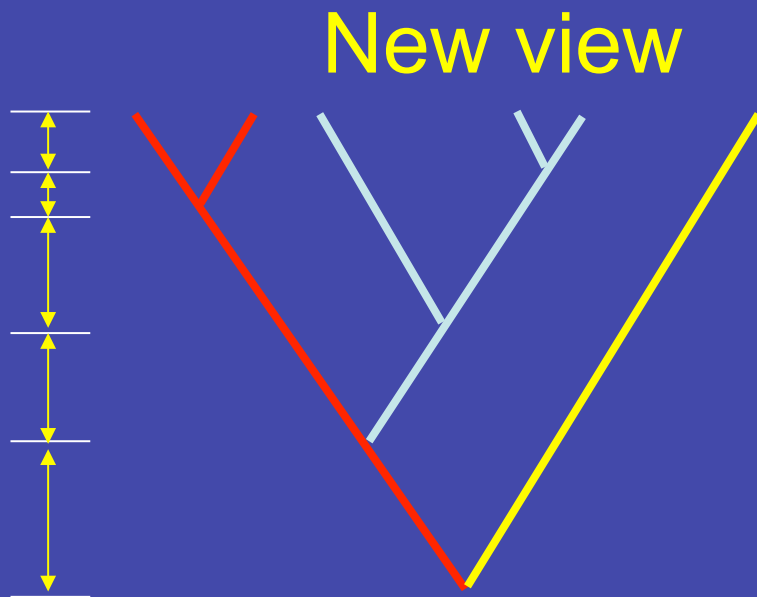
# The old population genetics

## Old view



Allele frequency change  
looking forward in time; alleles  
either the same or different

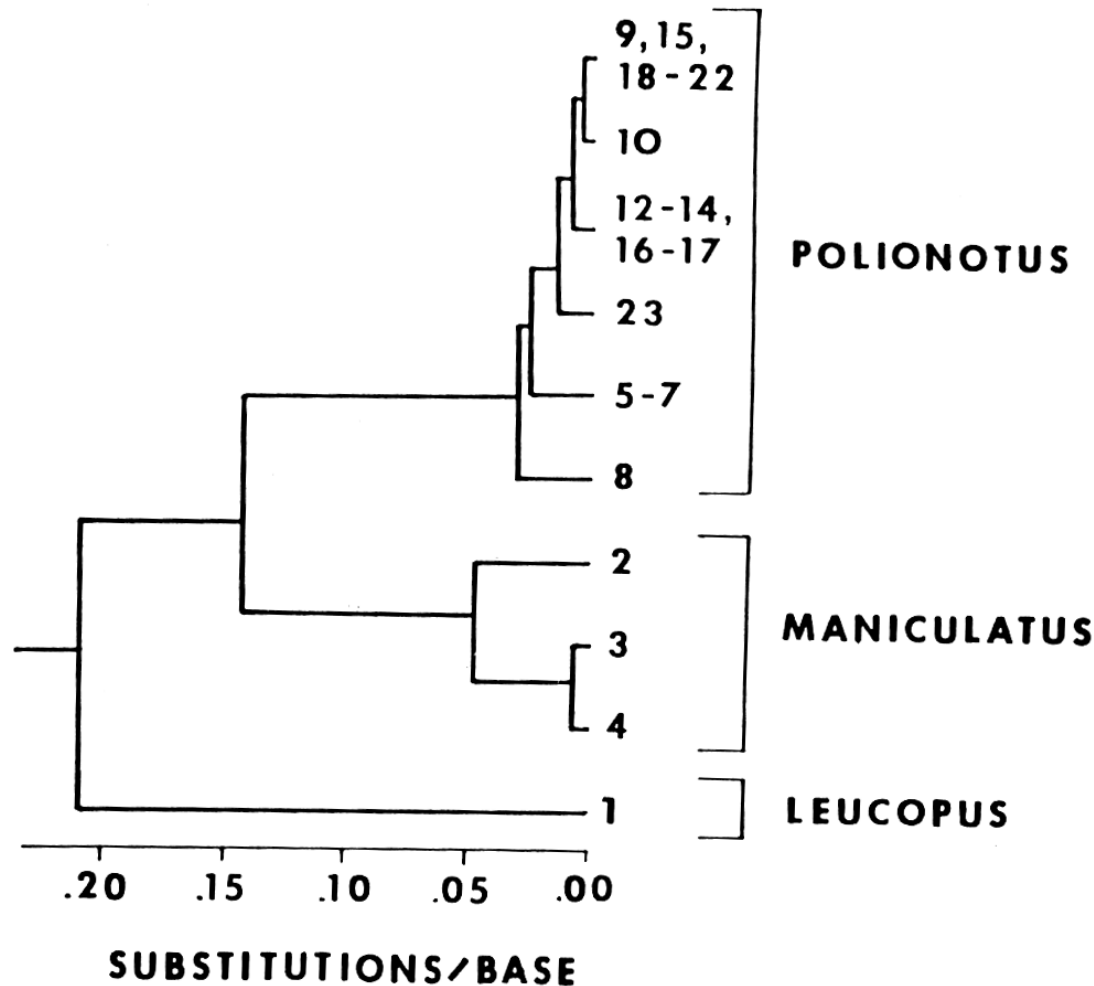
# The new population genetics



Shape of gene tree looking  
backward in time; alleles are  
related phylogenetically

# The first 'gene tree', 1979

J. C. AVISE, R. A. LANSMAN AND R. O. SHADE



Avise et al. (1979) *Genetics*

# Effective population size

- Abbreviated as  $N$  or  $N_e$
- The size of an ideal population with the same genetic dynamics as the population under study
- Dynamics = Genetic diversity, rate of loss of diversity, change in allele frequency over time
- Ideal = single population; all genetic variation neutral; no population structure or migration; no selection; random mating

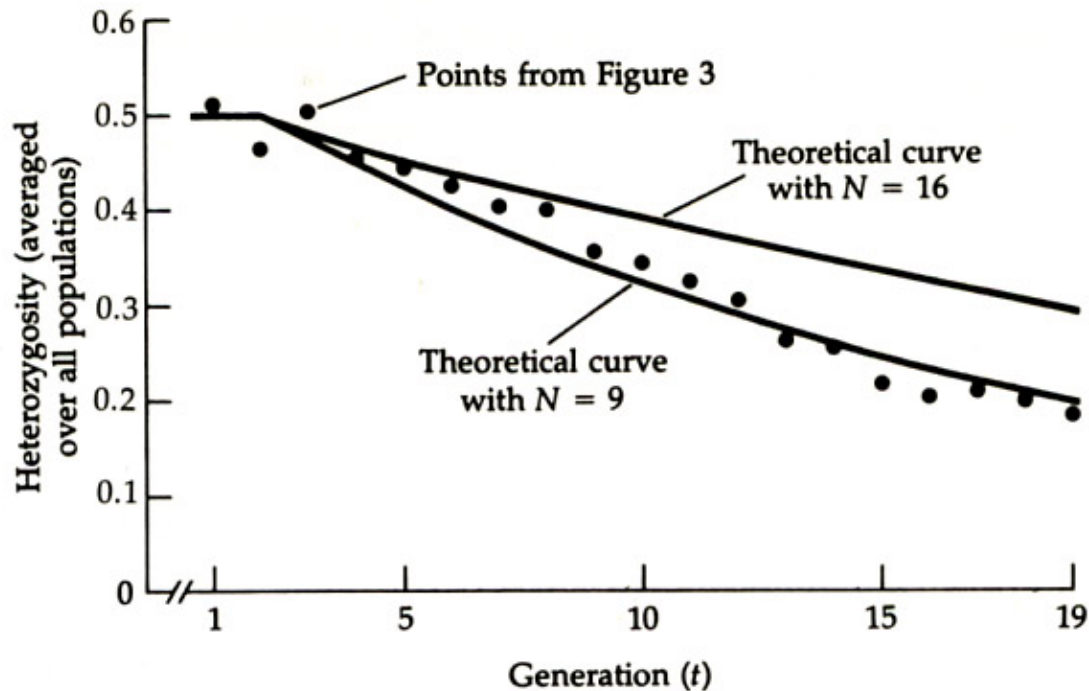


# Effective population size based on sex ratio



$$N_e = \frac{4N_M N_F}{N_M + N_F}$$

# “Loss of heterozygosity” effective population size



$$H = (0.5)(1 - 1/2N)^t$$

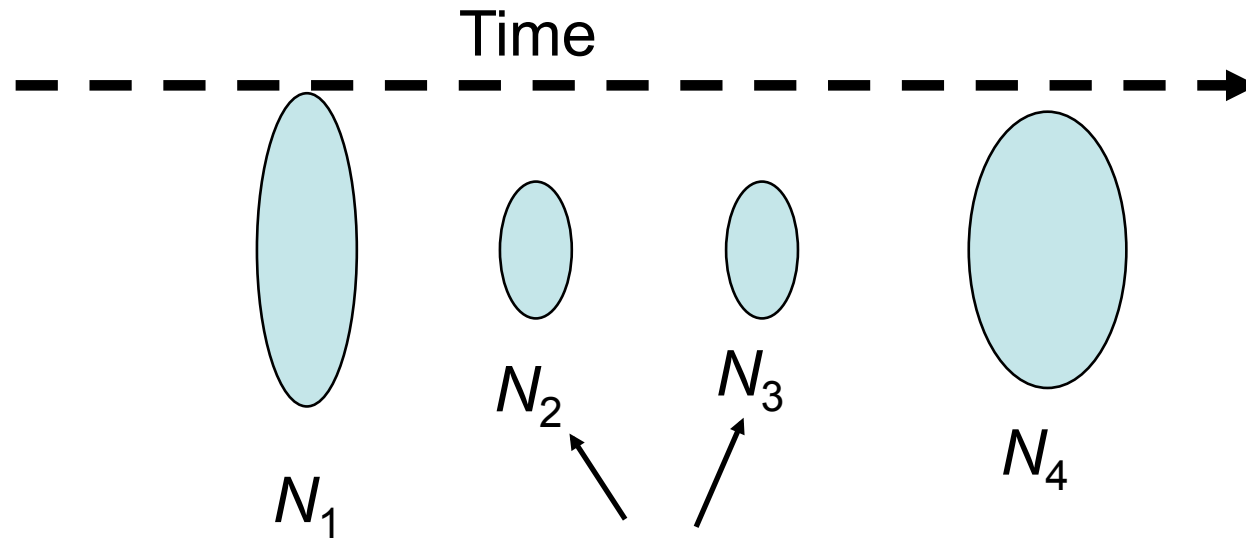
Heterozygosity  
(=variability) =  
proportion  
of individuals  
heterozygous  
at a locus

# Variance effective pop. size



$$N_e = \frac{4N - 2}{V_k + 2}$$

# Long-term effective population size as harmonic mean of temporal census sizes



$n = 4$   
populations

population bottlenecks

$$N_e = \frac{n}{1/N_1 + 1/N_2 + \cdots + 1/N_n}$$

# Estimating genetic diversity ( $4N\mu$ ) within populations

“Watterson’s theta”

$$\theta = \frac{S_n}{\sum_{i=1}^{n-1} \frac{1}{i}}$$

$S=3$

Allele 1: ACTGGCTGAACTT

Allele 2: ACTGGTTGAACTT

Allele 3: GCTGGTTGAACCT

\*
\*
\*

$\theta = 4N\mu$  = a function of the number of polymorphic sites in a population ( $S$ )

$n$  = number of alleles

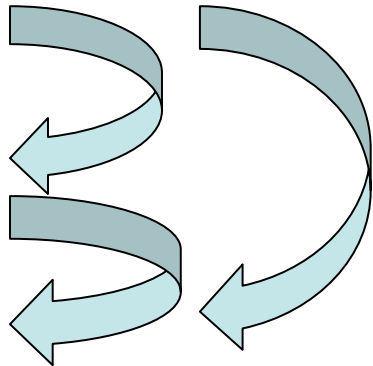
# A second estimate of $4N\mu$

$$\pi = \frac{\sum_{i=1}^n \sum_{j>1}^n k_{ij}}{\binom{n}{2}}$$

Allele 1: ACTGGCTGAACTT

Allele 2: ACTGGTTGAACTT

Allele 3: GCTGGTTGAACCT



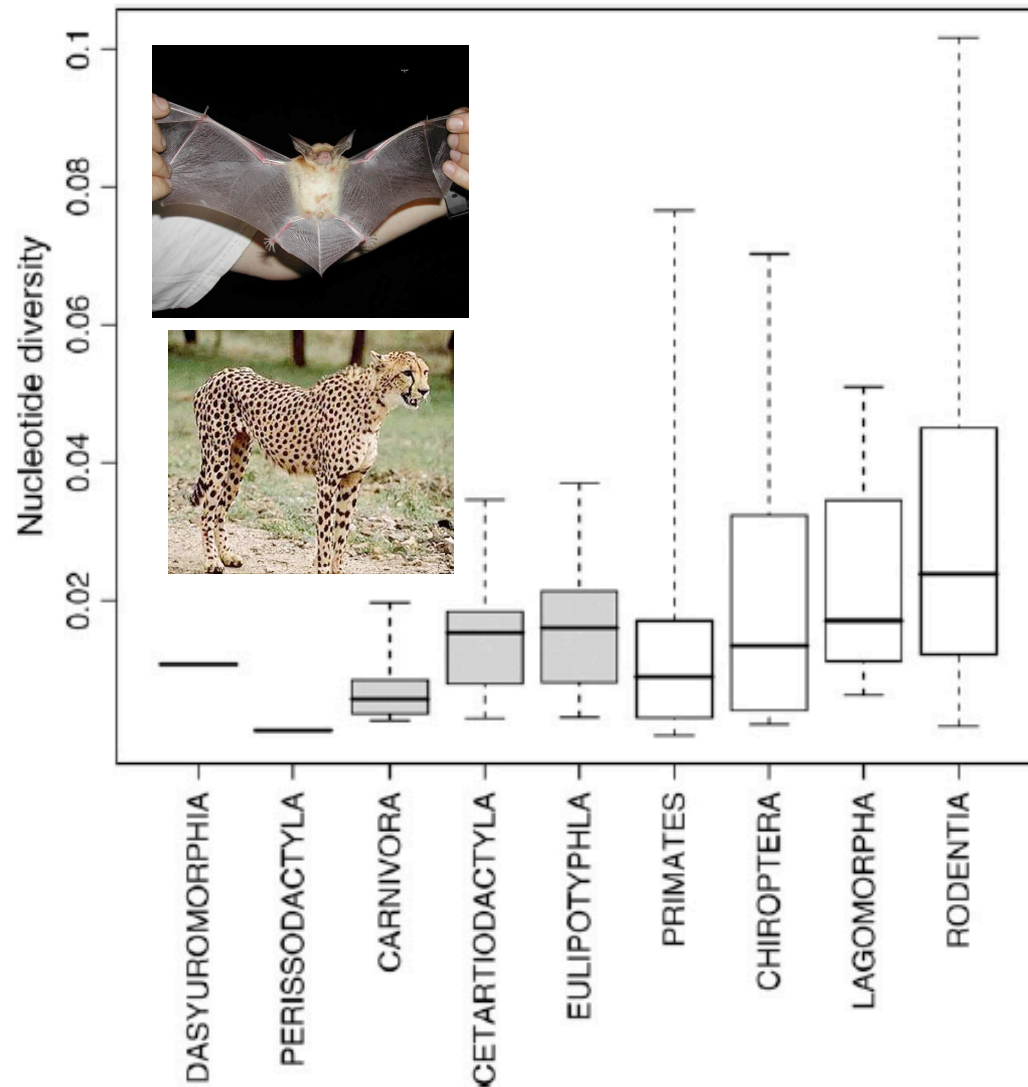
The diagram shows three alleles listed vertically. To the right of the alleles, there are three curved arrows forming a cycle: one from Allele 1 to Allele 2, one from Allele 2 to Allele 3, and one from Allele 3 back to Allele 1. The arrows are light blue with a darker blue outline.

$$\pi = \frac{k_{12} + k_{23} + k_{13}}{\binom{3}{2}} = \frac{k_{12} + k_{23} + k_{13}}{3}$$

$k$  = number of differences between pairs of alleles  $i$  and  $j$

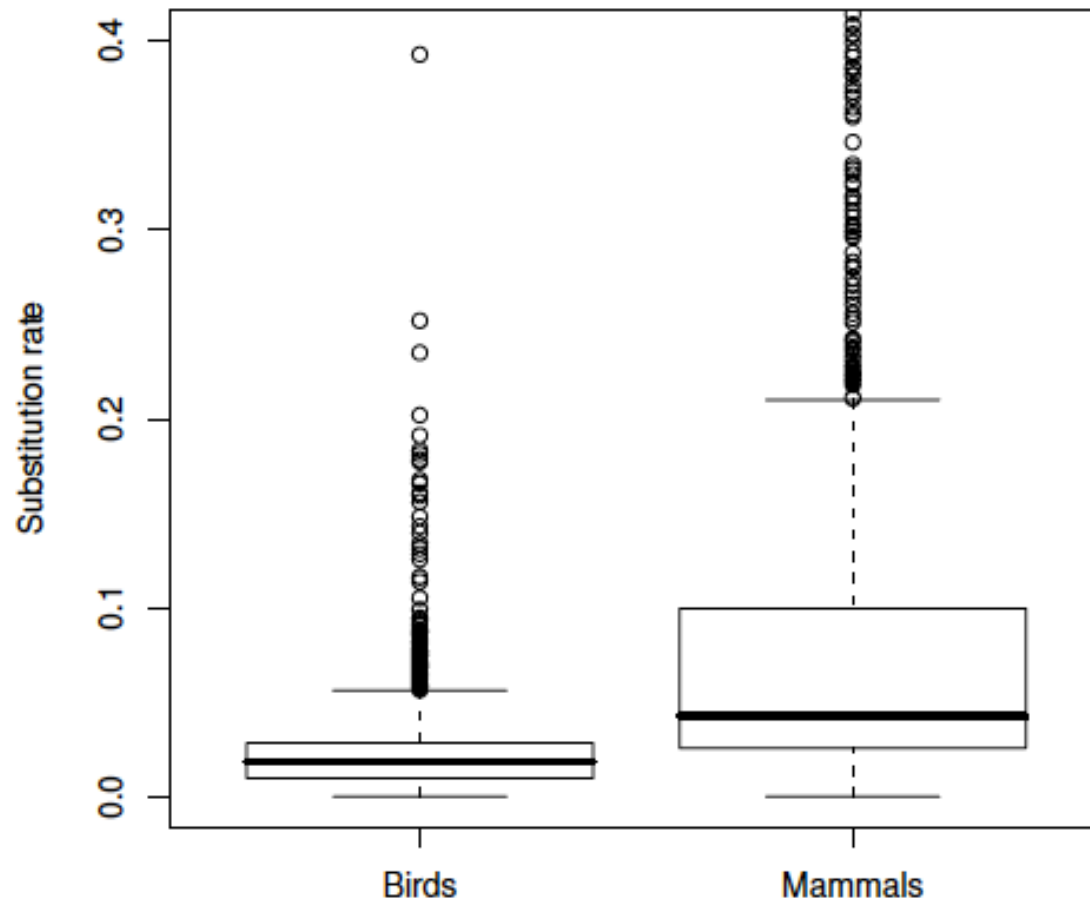
$n$  = number of alleles

# Nucleotide diversity in mammals



Nabholz et al. *Genetics* 178: 351–361

# Determinants of nucleotide diversity in birds



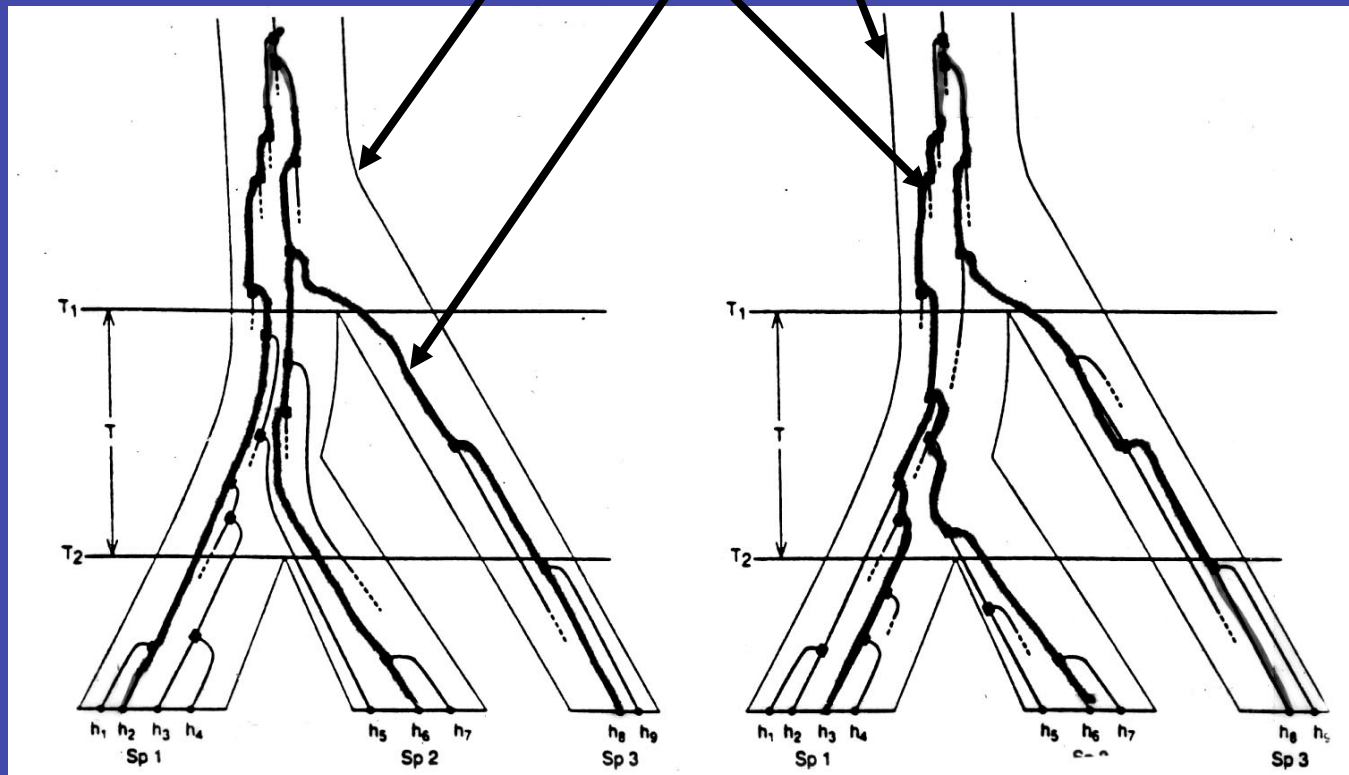


# Two rules of gene trees near the species boundary

1. Gene trees don't always match the species tree
2. Gene divergence often precedes population divergence

Species tree

Gene tree

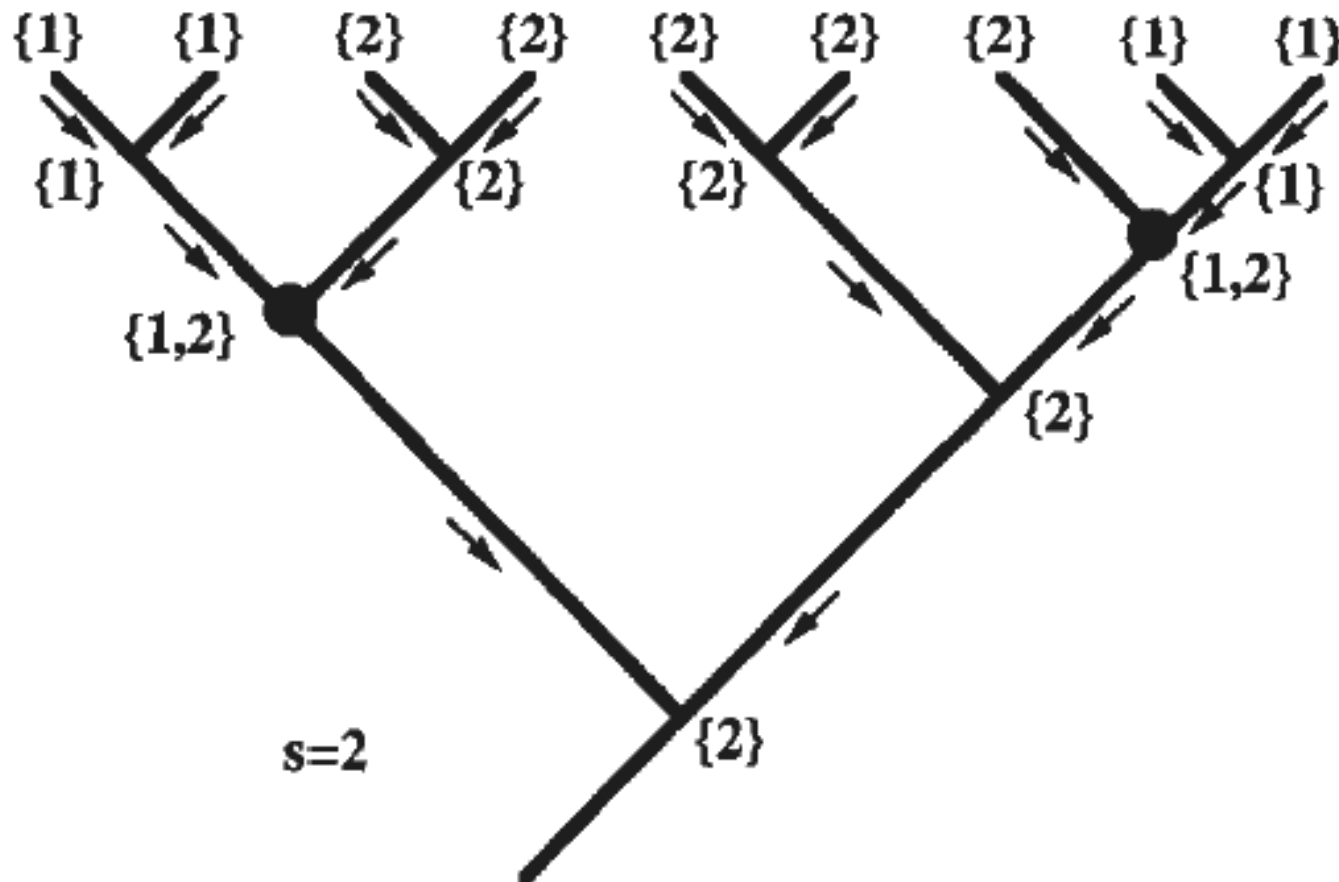


$$\begin{array}{c} \uparrow \\ 2N_e \\ \downarrow \\ \uparrow \\ T/4N_e \\ \downarrow \end{array}$$

Incomplete lineage sorting

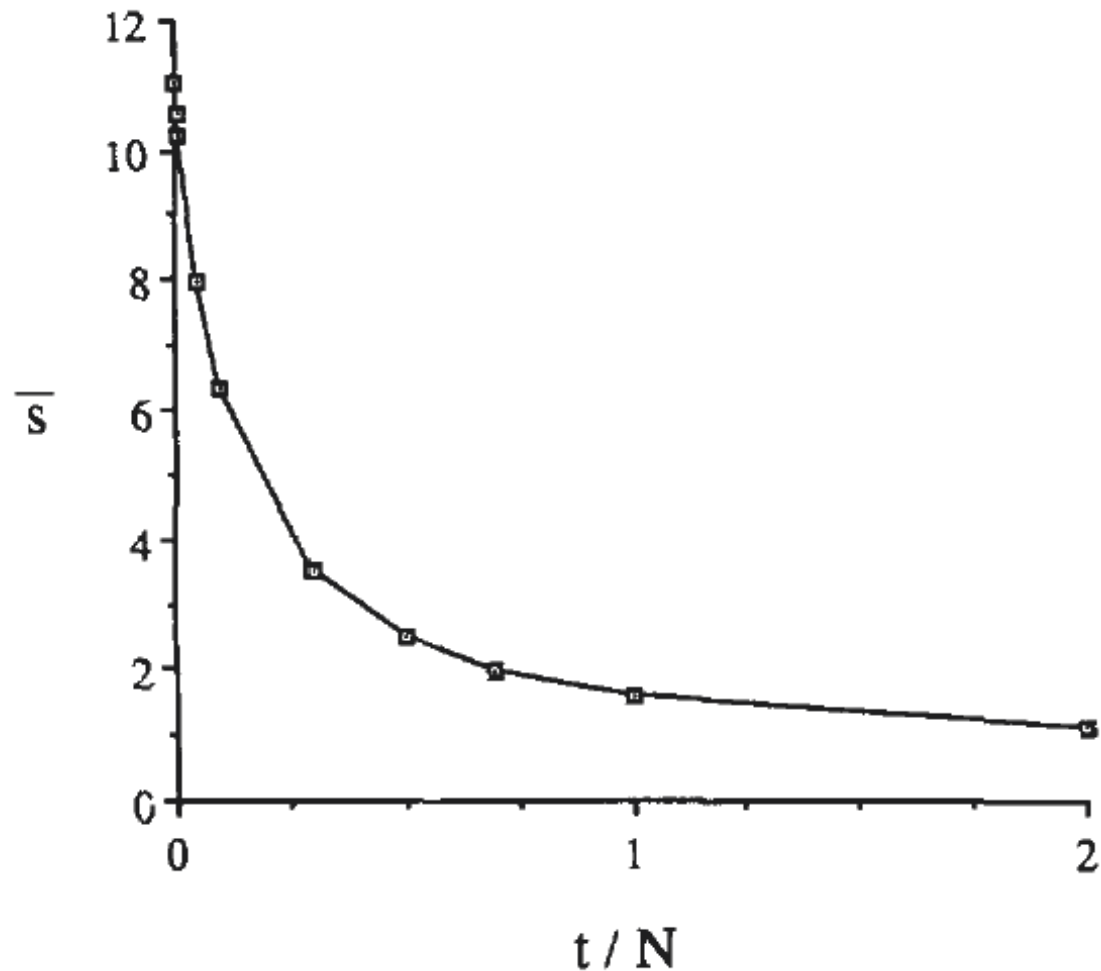
Hudson 1990. Oxf. Surv. Evol. Biol. 7: 1-44.

# Counting the number of interpopulation coalescent events



Slatkin and Maddison. 1989. Genetics 123: 603-613

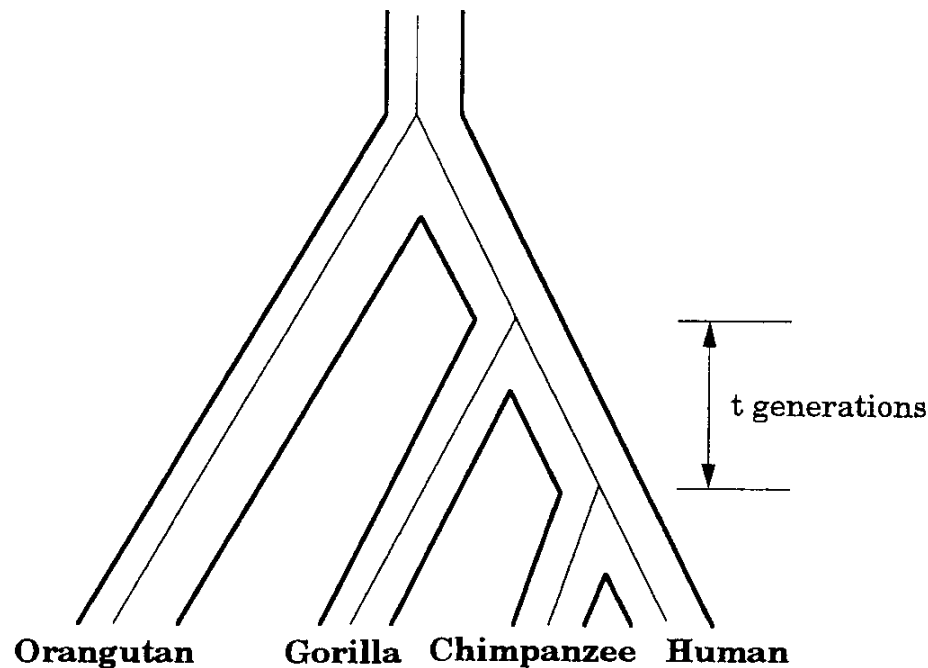
# $s$ as an index of divergence time



Slatkin and Maddison. 1989. Genetics 123: 603-613

# Gene trees and species trees in primates

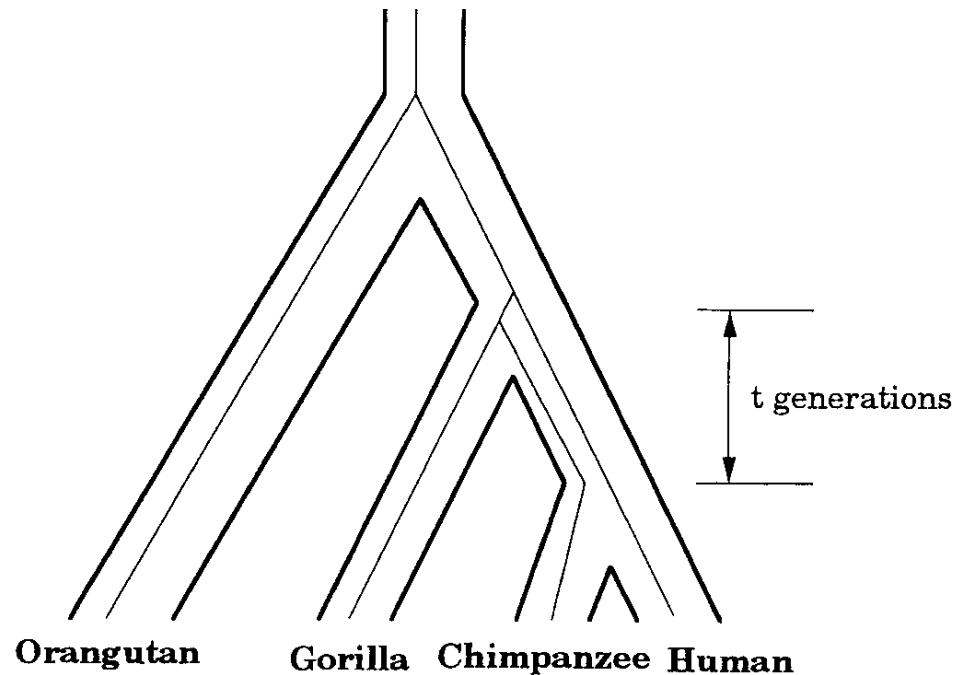
Gene tree compatible  
to species tree



Satta et al. 2000. Molecular Phylogenetic  
s and Evolution 14: 259–275.

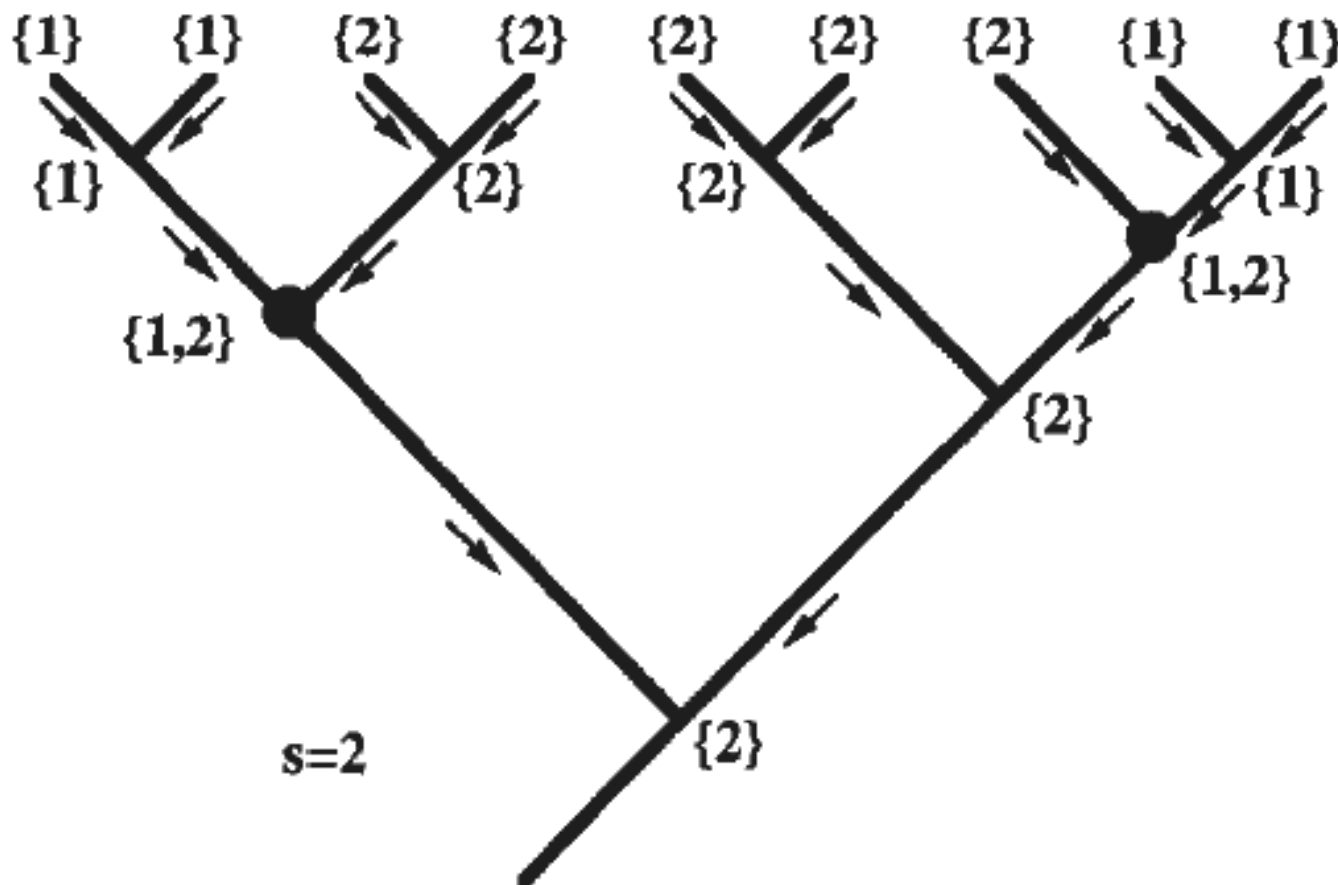
# Gene trees and species trees in primates

Gene tree incompatible  
to species tree



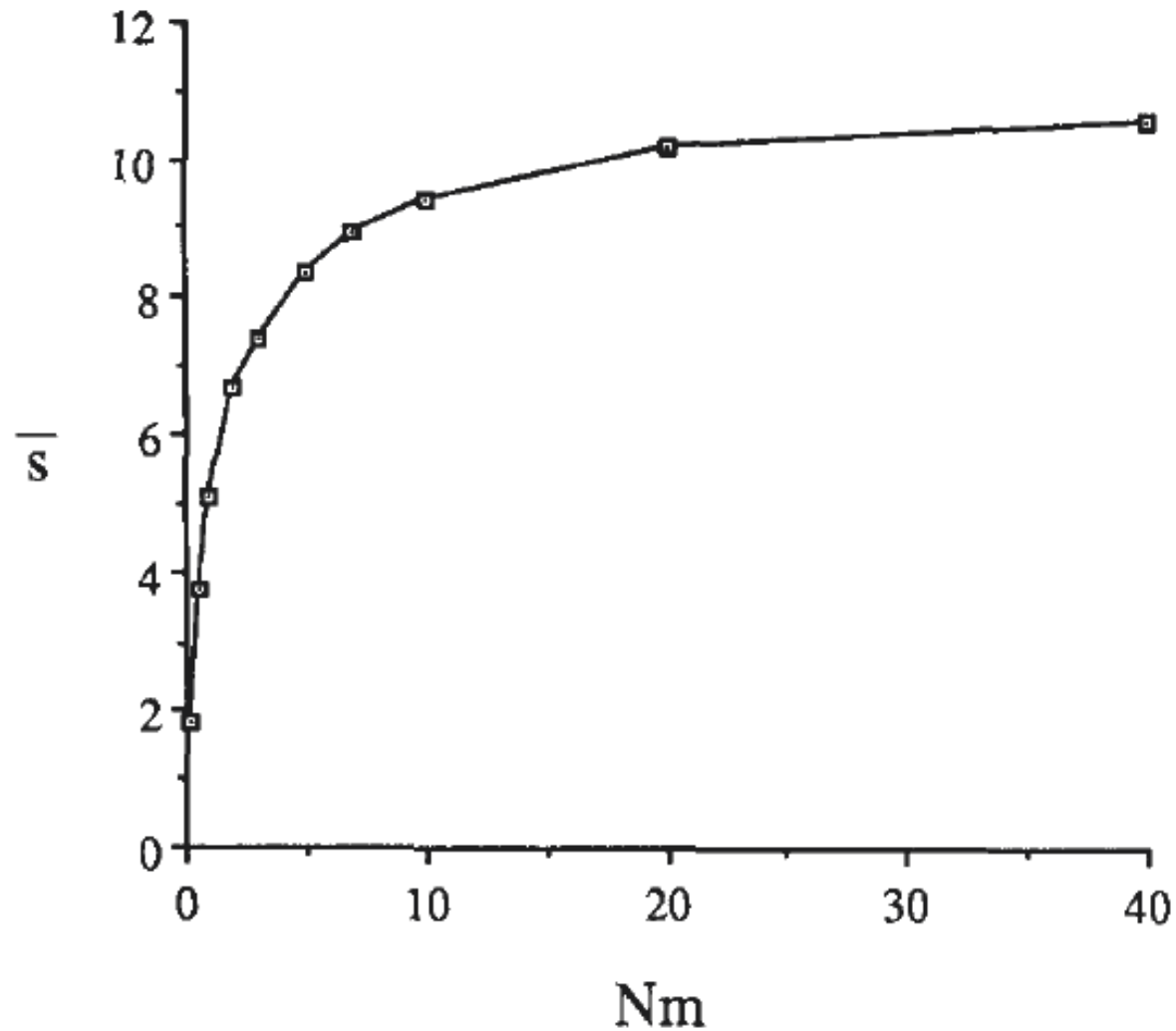
Satta et al. 2000. Molecular Phylogenetics  
and Evolution 14: 259–275.

# Counting the number of interpopulation coalescent events



Slatkin and Maddison. 1989. Genetics 123: 603-613

$s$  as an index of gene flow

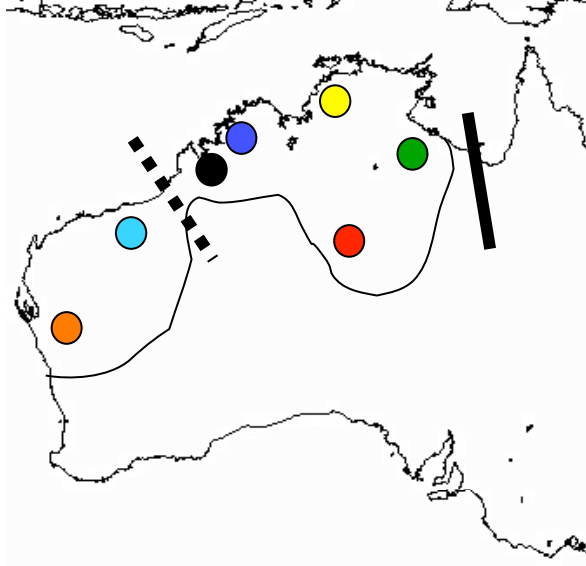




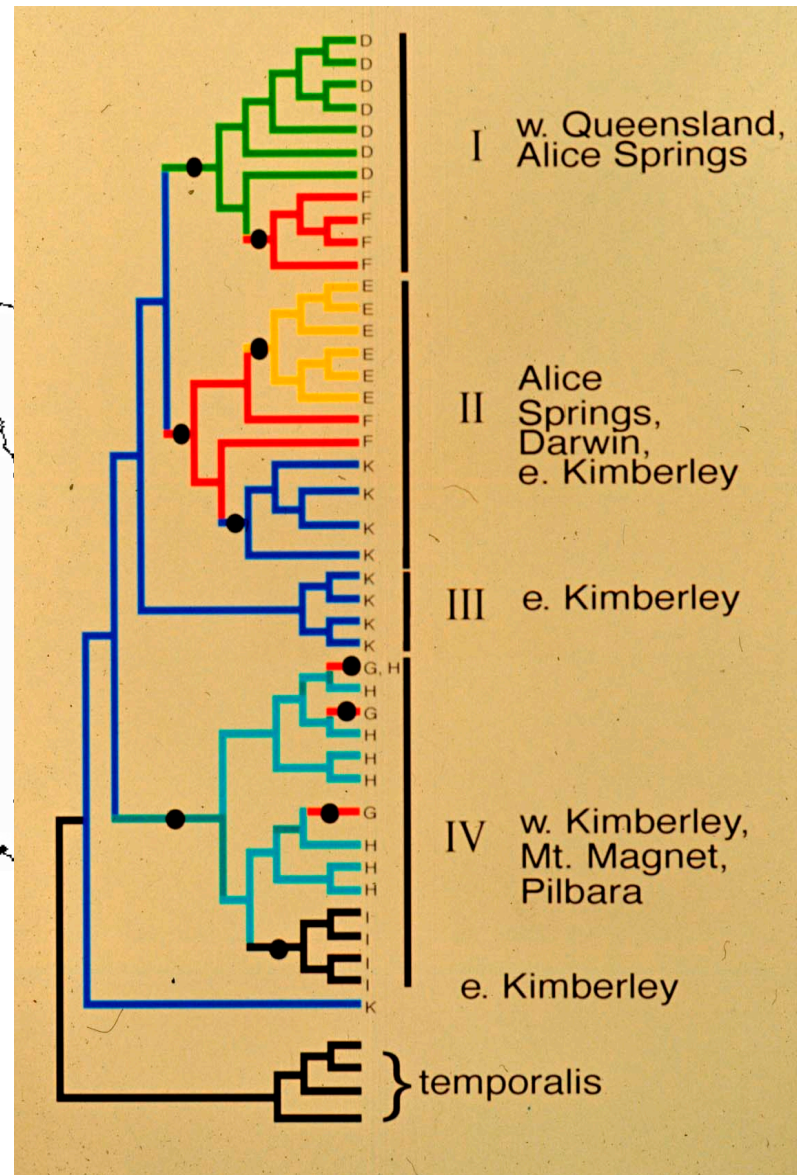
# Gene flow erodes population monophyly



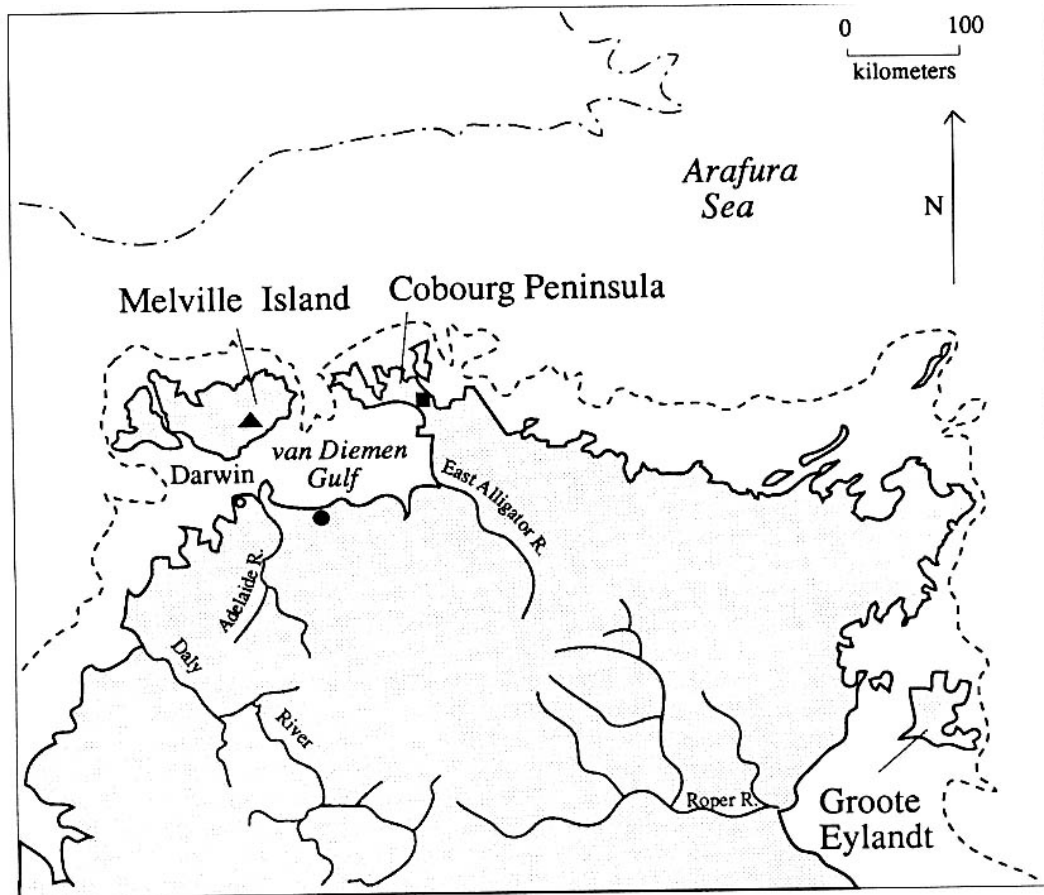
Photo: Graeme Chapman



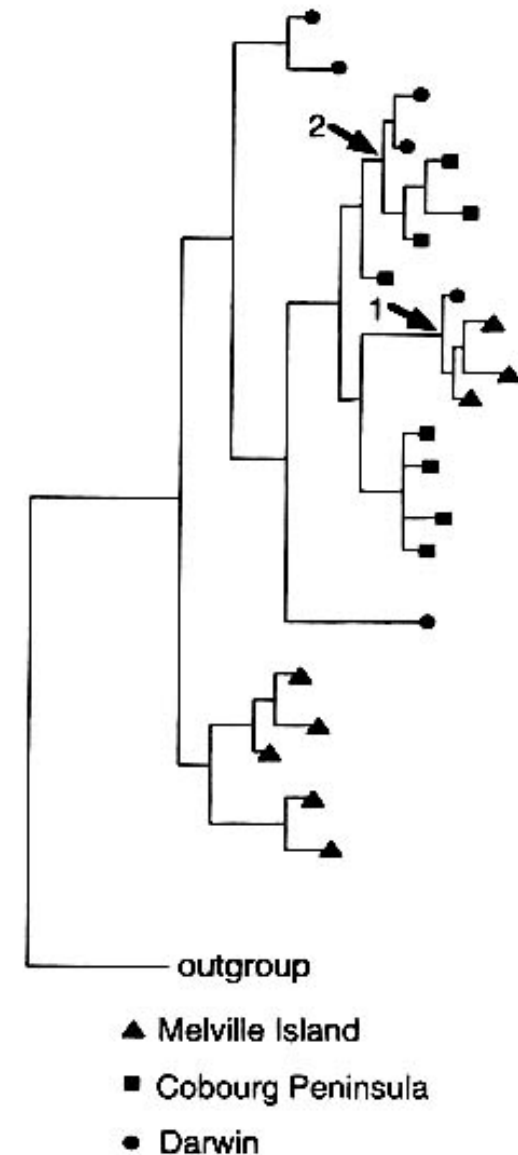
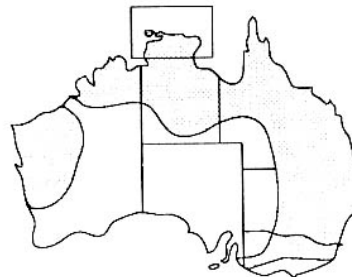
- Migration event reconstructed by parsimony



# Gene flow erodes population monophyly



- Shoreline approximately 18,000 years B.P.
- .... Shoreline approximately 10,000 years B.P.



# Genetic differentiation between populations

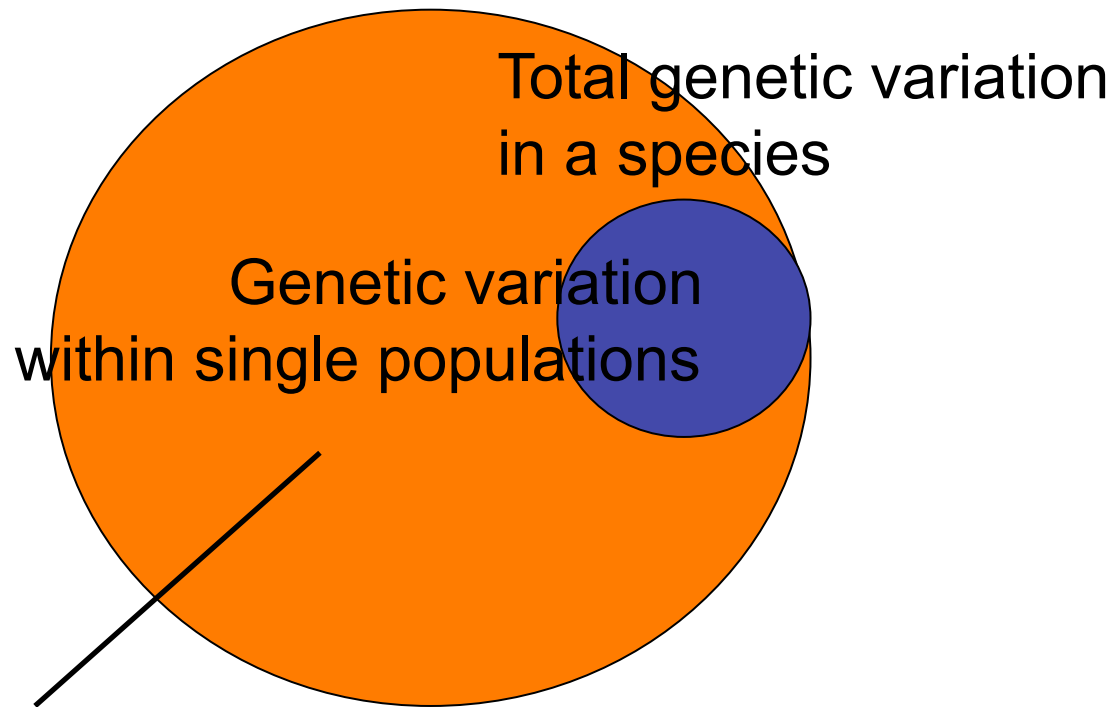
$$F_{st} = \frac{\theta_b - \theta_w}{\theta_b}$$

or

$$F_{st} = \frac{\pi_b - \pi_w}{\pi_b}$$

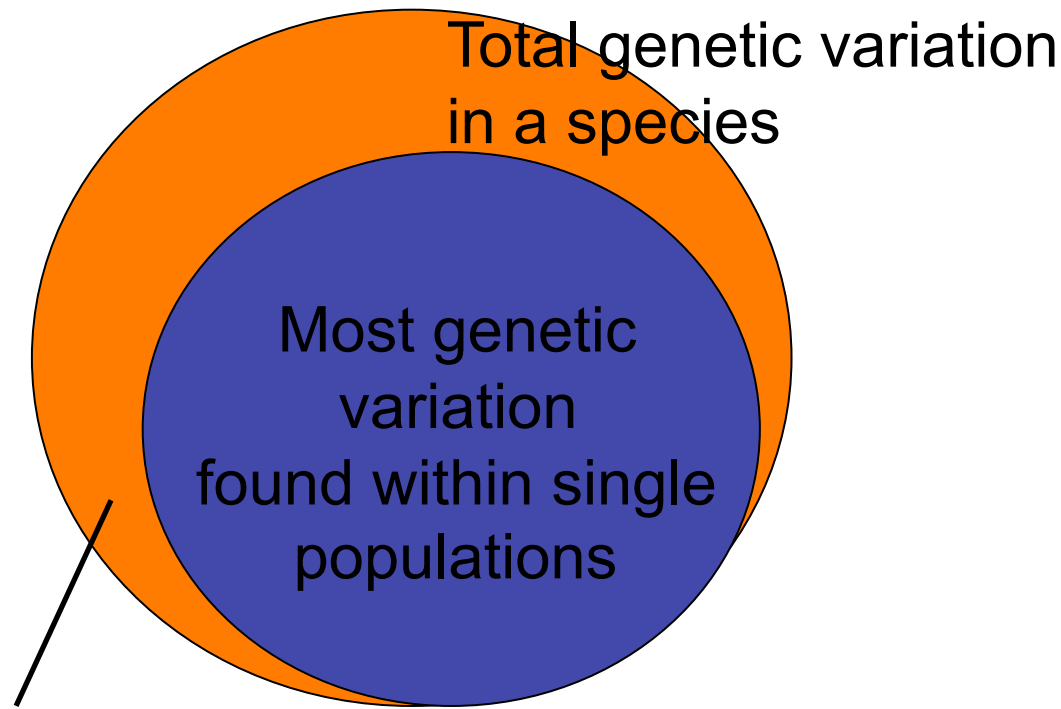
$b$ , between populations;  $w$ , within populations

# Genetic differentiation between populations



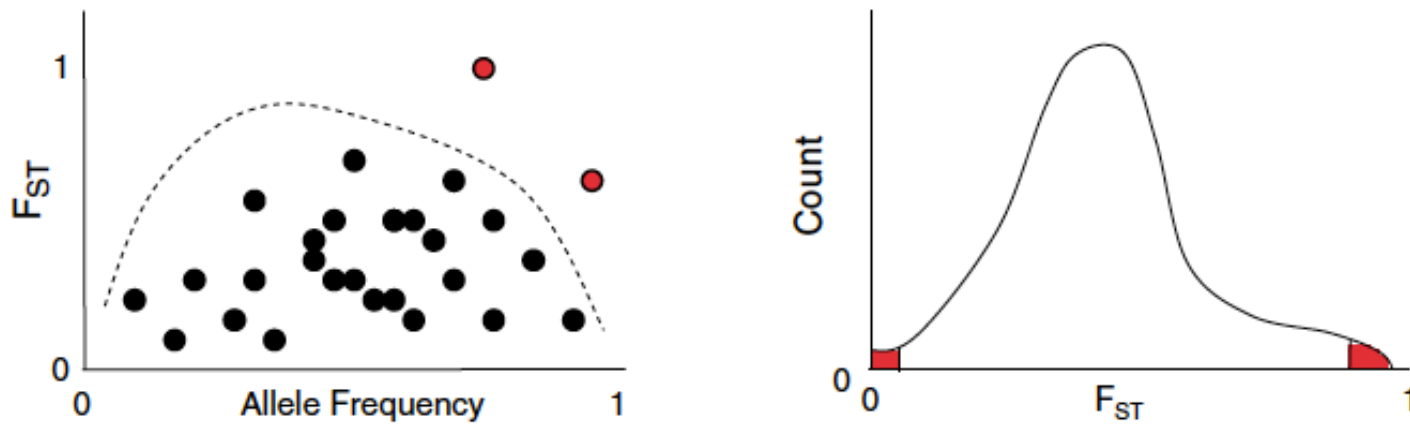
Most genetic variation  
found between populations:  $F_{st}$  high

# Genetic differentiation between populations

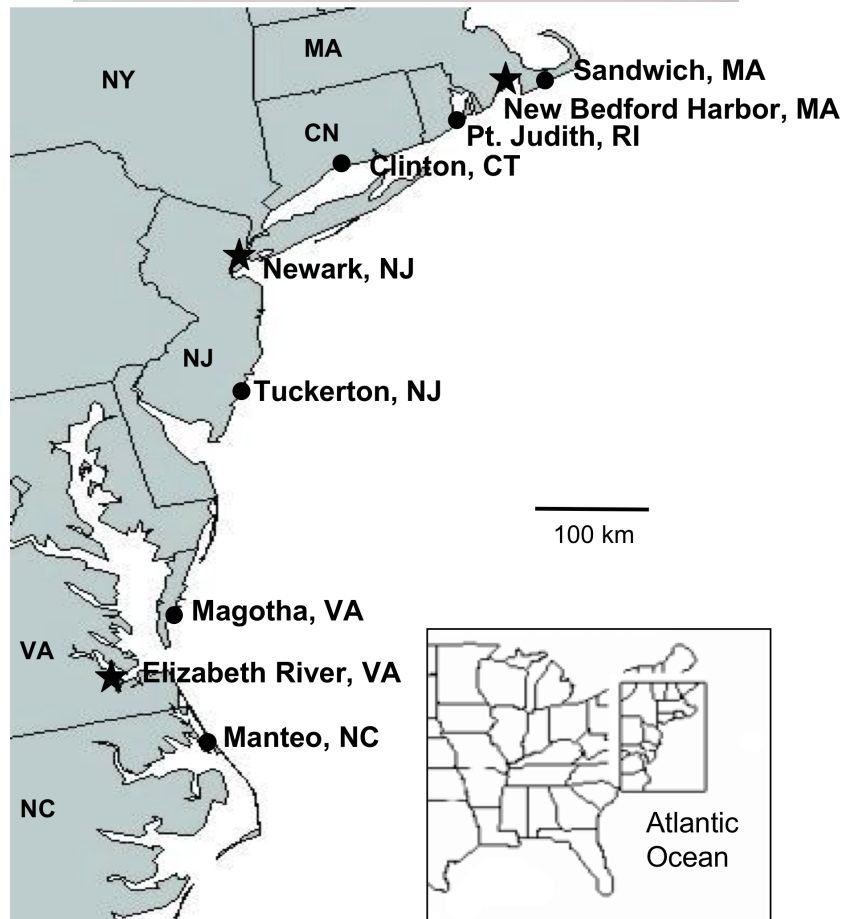


Little genetic variation  
found between populations:  $F_{st}$  low

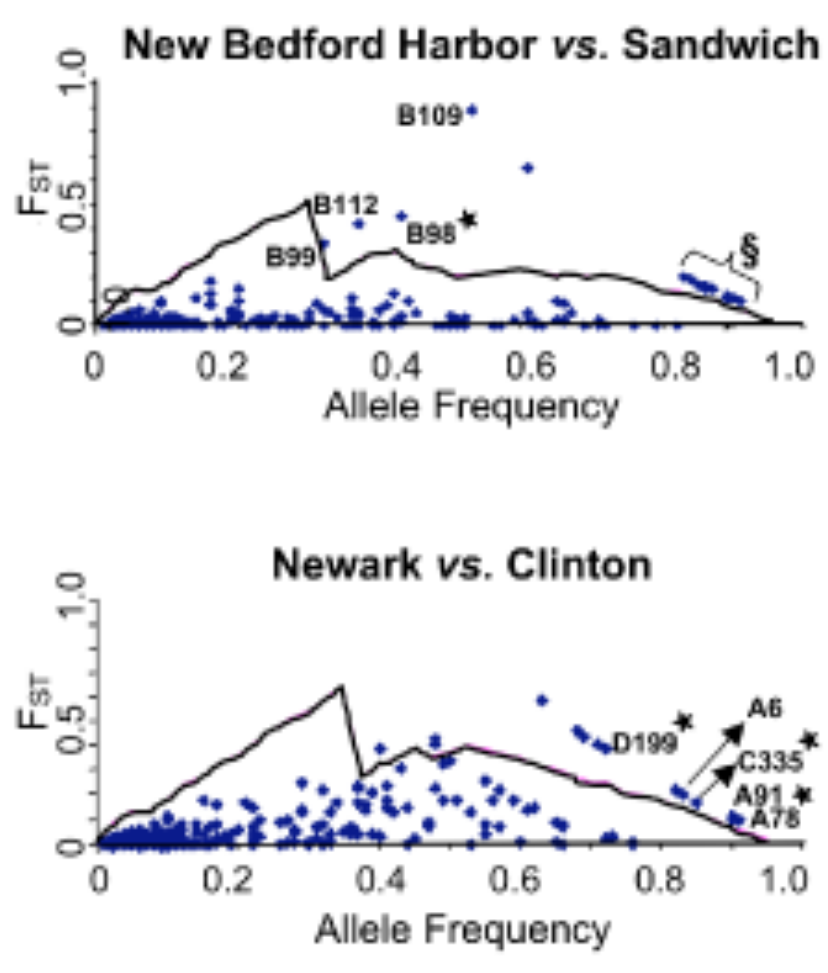
# Identifying outlier loci using $F_{st}$



# Identifying loci under pollution-driven selection using $F_{st}$ and outlier loci



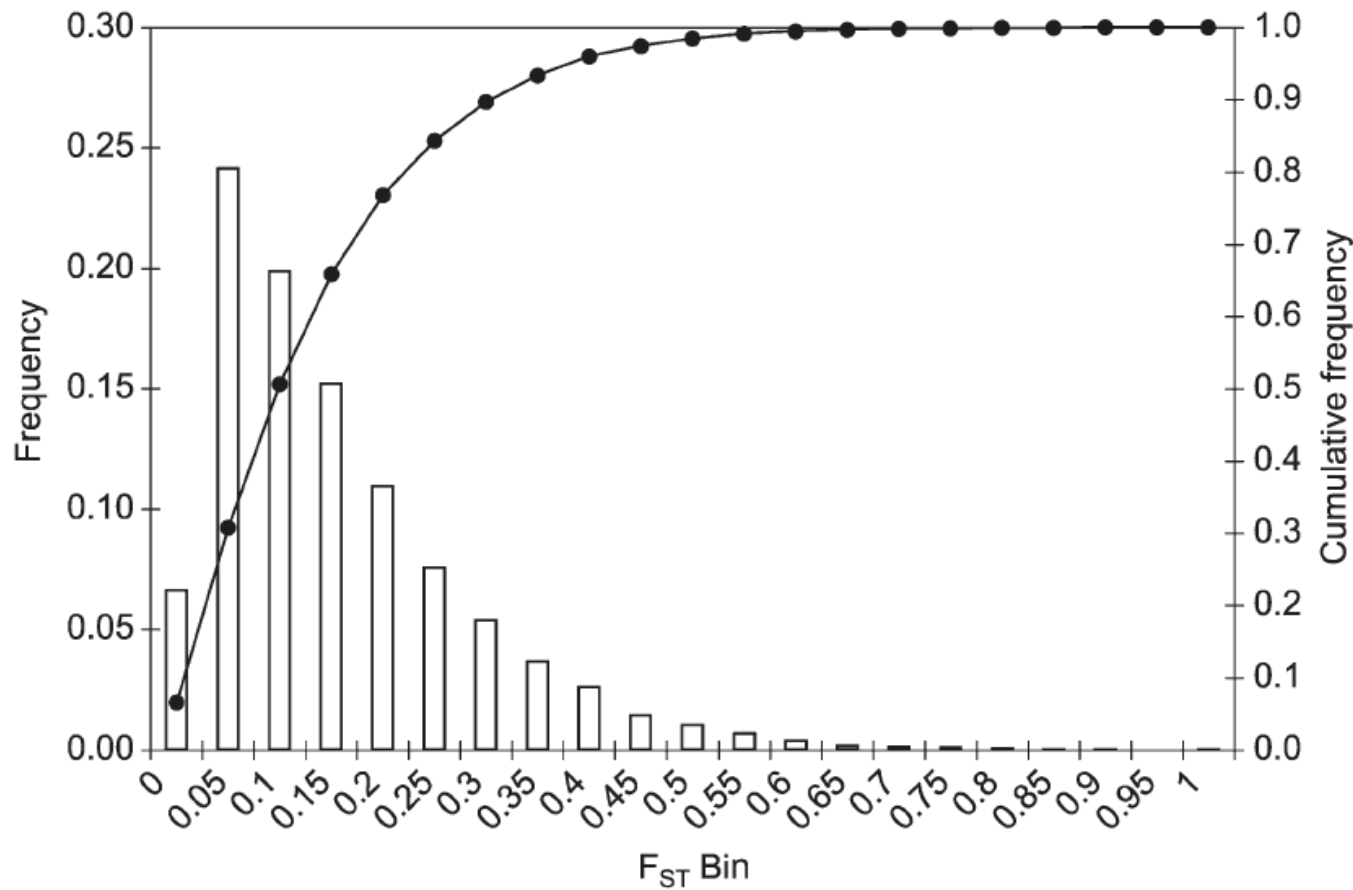
# Identifying loci under pollution-driven selection using $F_{st}$ and outlier loci



Whitehead et al. 2010. *Molecular Ecology* 19:5186-5203

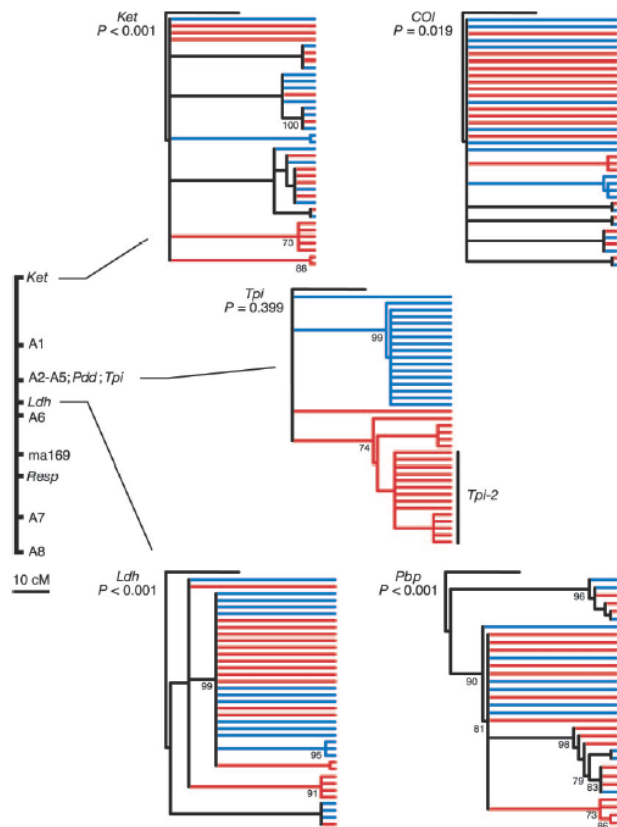


# Distribution of $F_{st}$ among 8525 loci in humans



Shriver et al. (2004) *Human Genomics* 1: 274–286.

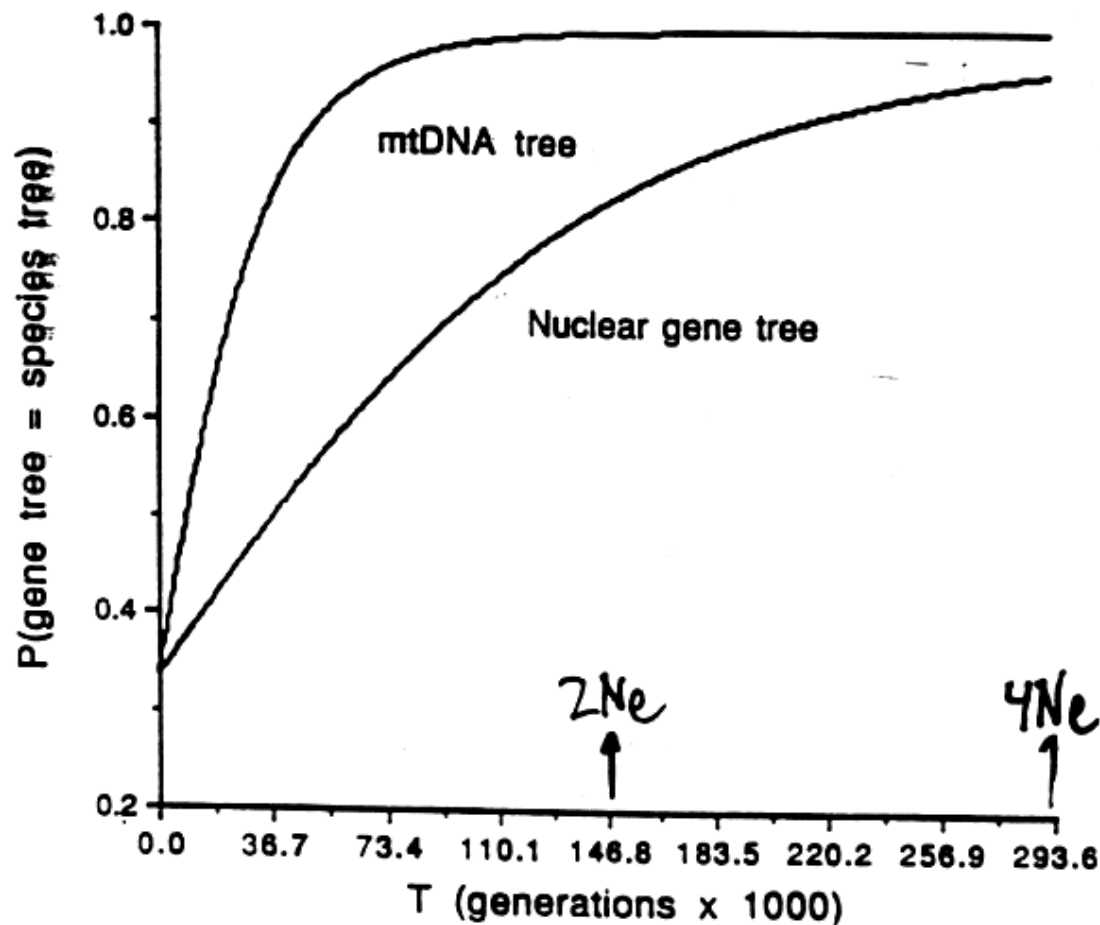
# Gene tree monophyly as an indicator of natural selection



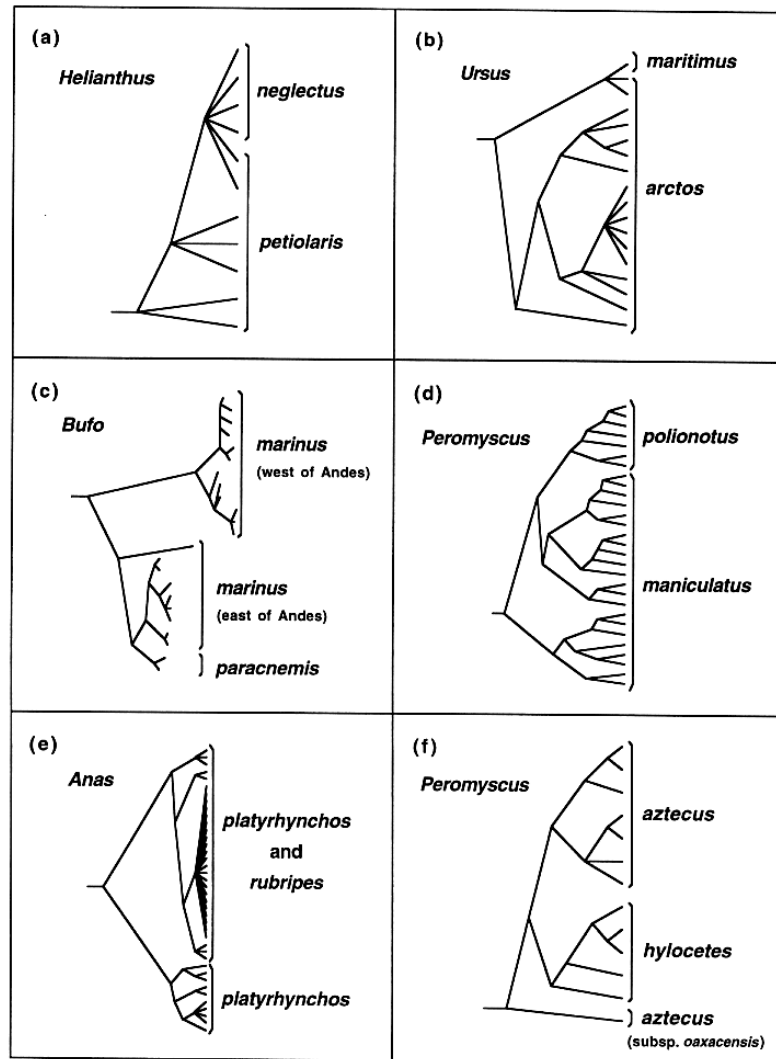
European corn borer

Dopman et al. 2005. PNAS 102: 14706-14711

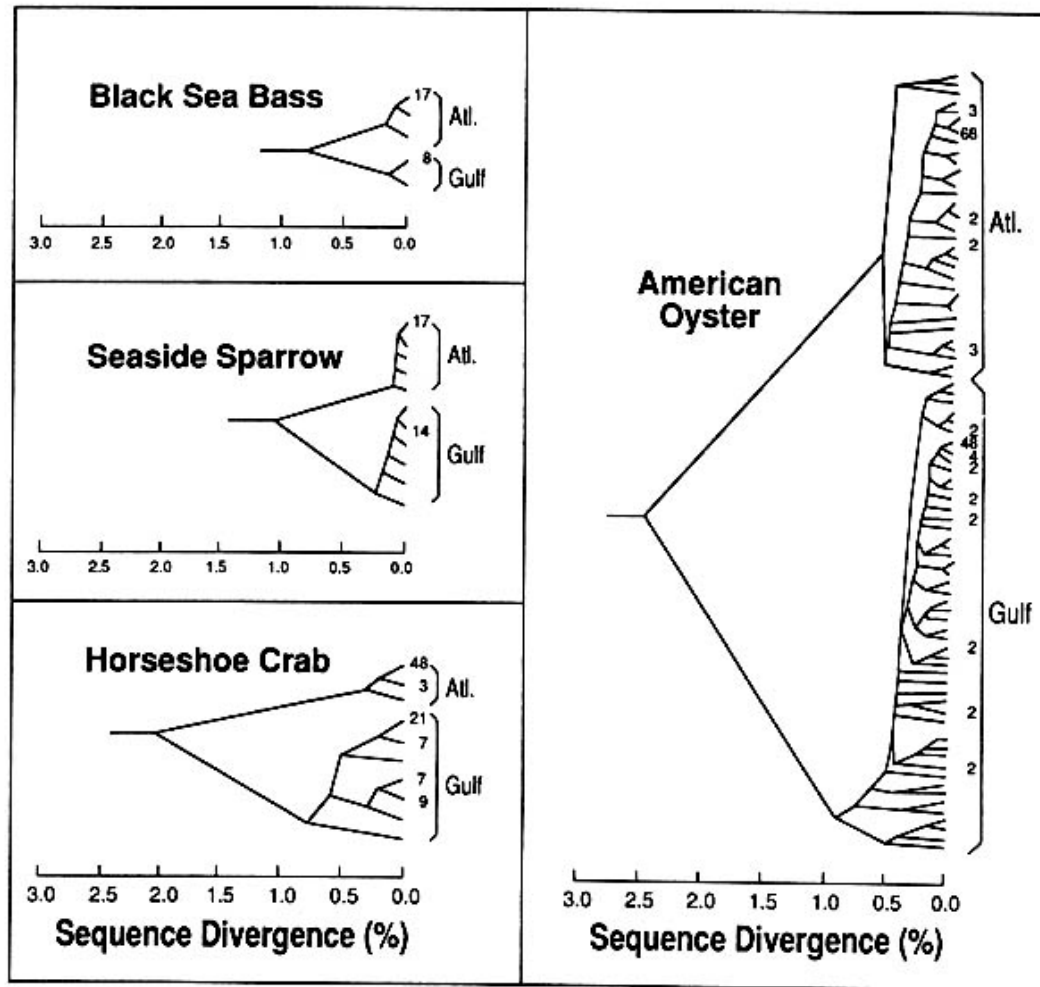
# Faster approach to concordance for mtDNA versus nuclear DNA



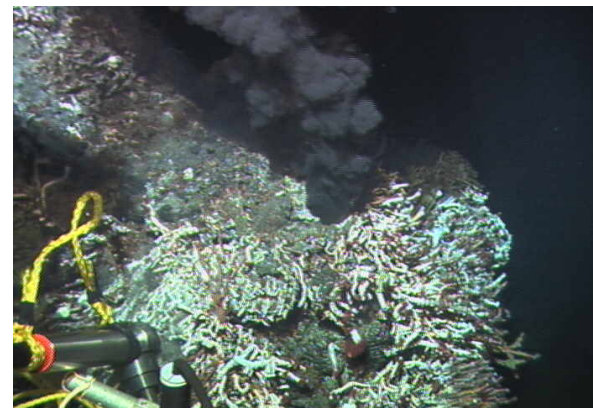
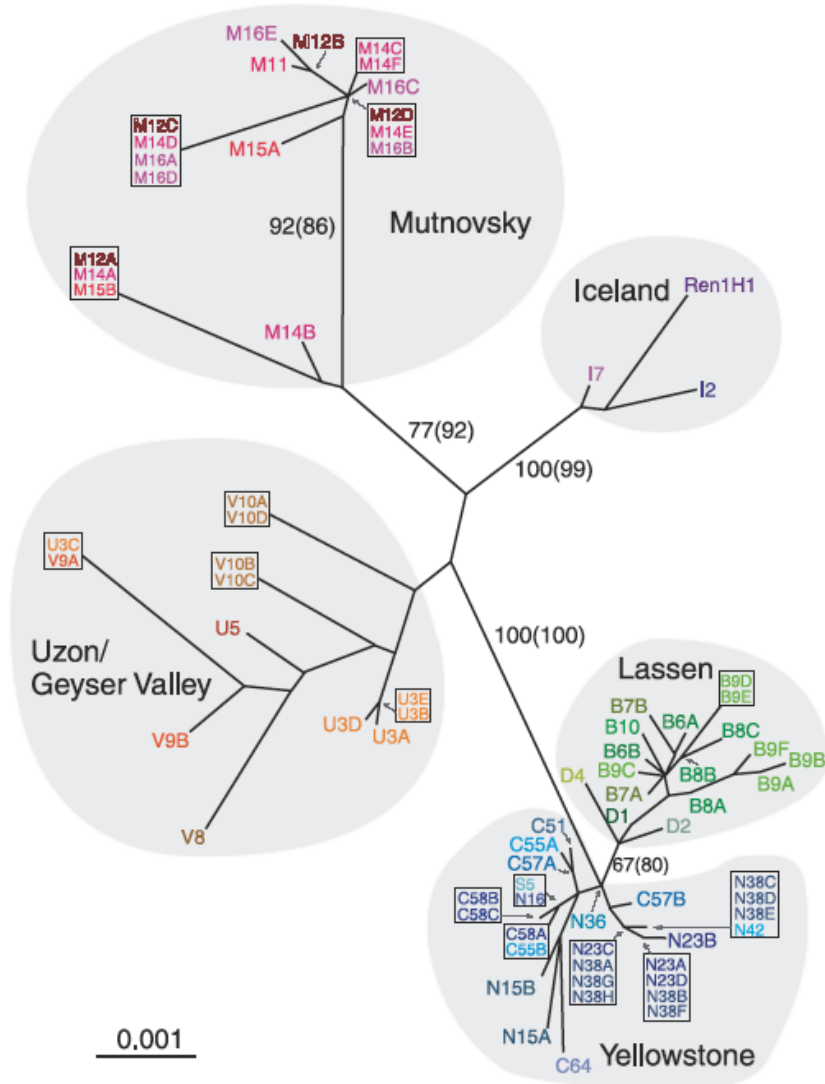
# Examples of discordance between gene and species trees



# Examples of large mtDNA breaks within species

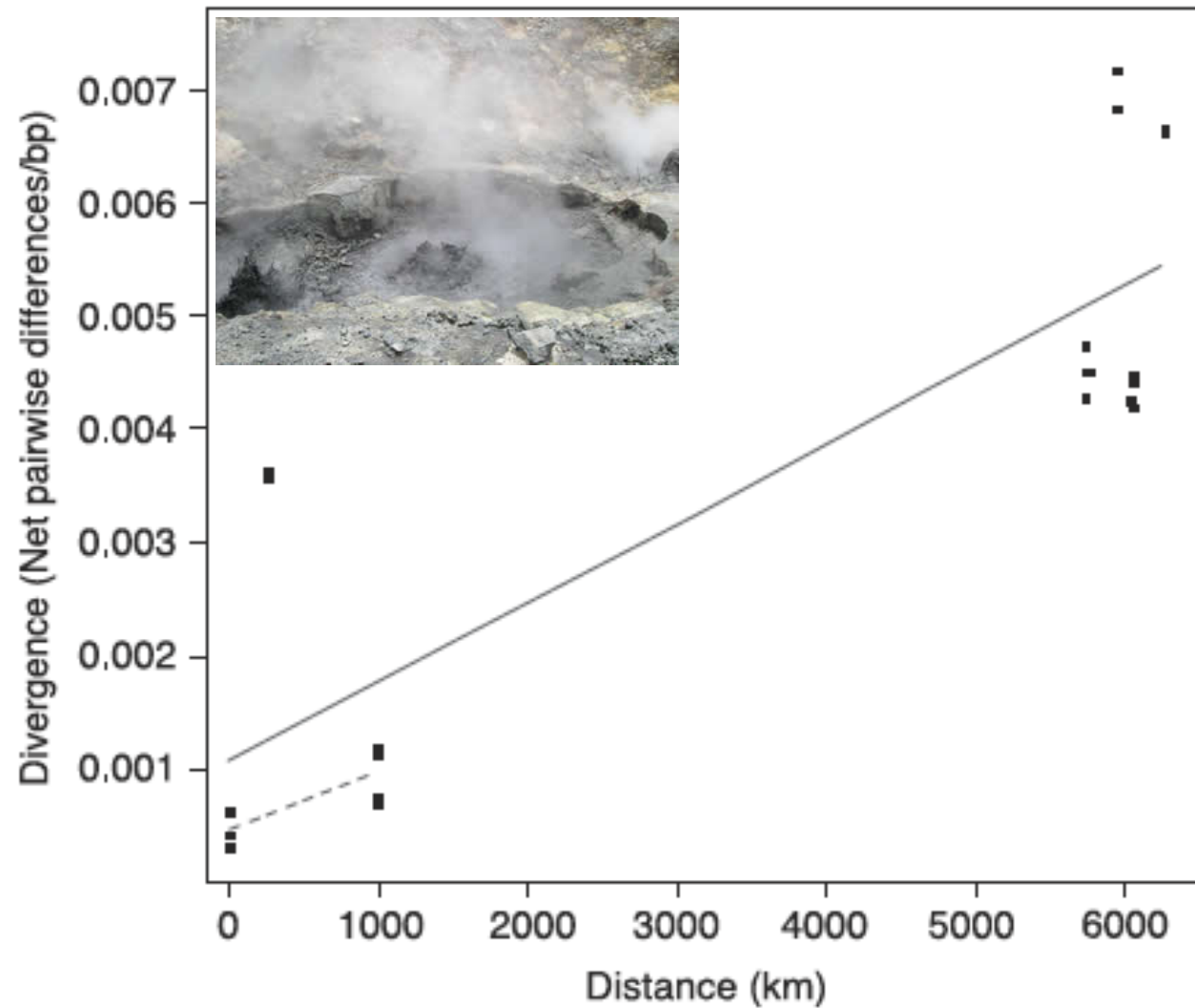


# Structured populations of hyperthermophilic Archaea

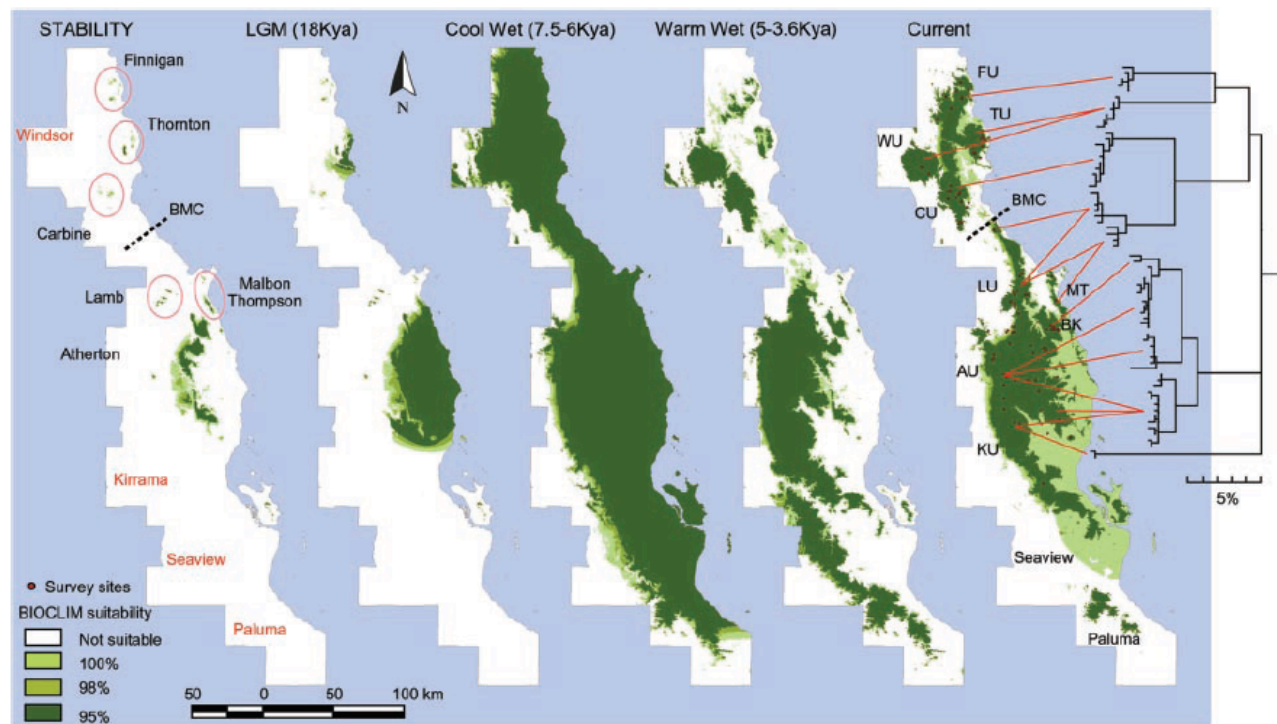




# Isolation by distance



# Genetic diversity and climate stability



Hugall, A., C. et al. *PNAS* 99:6112-6117.