

Everything you wanted to know about Frogs



Everything you wanted to know about ~~mainframe~~ Computers



Some things I wanted to know about Population genetics

$$P(S|G, E) = \frac{\prod_j \frac{k_j(k_j-1)}{\theta_j} \rho_j}{\prod_j \frac{k_j(k_j-1)}{\theta_j} \rho_j + \sum_i \frac{\sum_j k_i M_{ji} (1-\rho_j)}{\theta_i}} + \frac{\sum_i \frac{\sum_j k_i M_{ji} (1-\rho_j)}{\theta_i}}{\prod_j \frac{k_j(k_j-1)}{\theta_j} \rho_j + \sum_i \frac{\sum_j k_i M_{ji} (1-\rho_j)}{\theta_i}}$$

so we can assemble all parts to get the final result

there are $\frac{2}{k_i(k_i-1)}$ per coalescent events on interval and M_{ji} migration $\frac{1}{\theta_i}$ per time interval

$$\exp\left(-u \left(\sum_i \frac{k_i(k_i-1)}{\theta_i} + \sum_i \sum_j k_i M_{ji} \right) \left(\sum_i \frac{k_i(k_i-1)}{\theta_i} + \sum_i \sum_j k_i M_{ji} \right) \right) * \begin{cases} 1 & \text{if coalescence} \\ \rho & \text{if coalescence, otherwise} \end{cases}$$

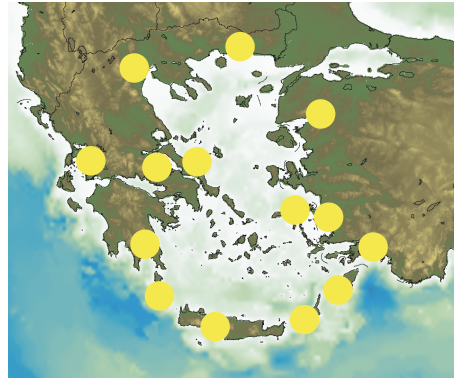
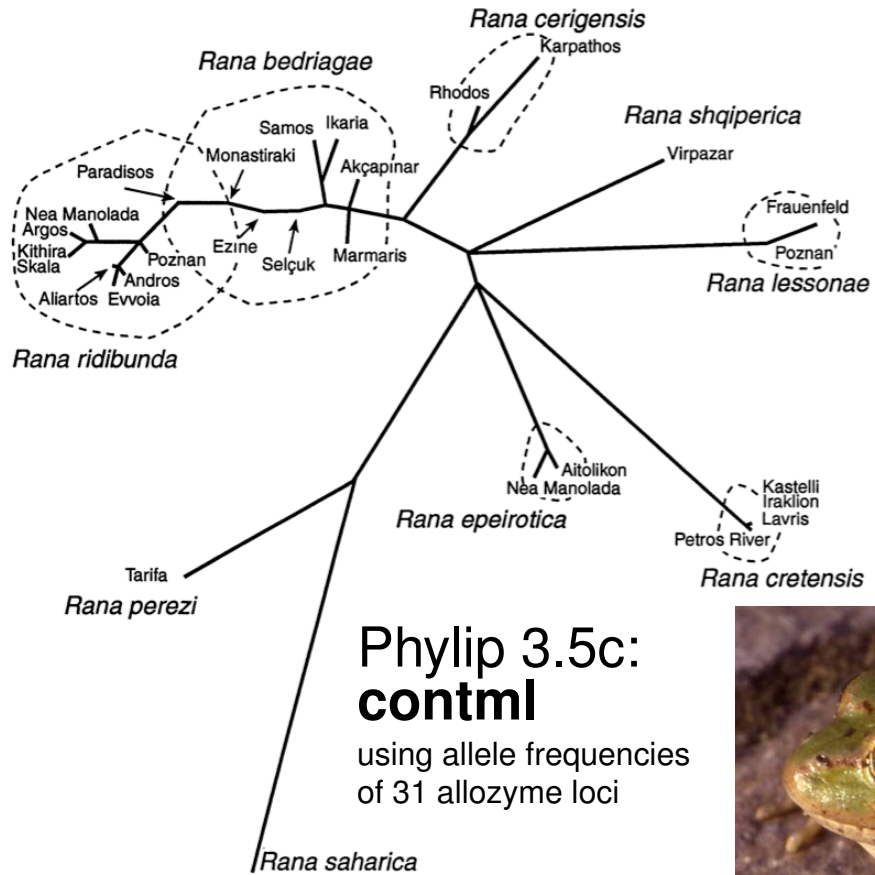
$$\left(\frac{\prod_i \frac{k_i(k_i-1)}{\theta_i} \rho_i + \sum_j \frac{\sum_i k_i M_{ji} (1-\rho_j)}{\theta_i}}{\prod_i \frac{k_i(k_i-1)}{\theta_i} \rho_i + \sum_j \frac{\sum_i k_i M_{ji} (1-\rho_j)}{\theta_i}} \right) * \left(\frac{\frac{k_j(k_j-1) \rho_j}{\theta_j} + \sum_i \frac{k_i M_{ji} (1-\rho_j)}{\theta_i}}{\prod_i \frac{k_i(k_i-1)}{\theta_i} \rho_i + \sum_j \frac{\sum_i k_i M_{ji} (1-\rho_j)}{\theta_i}} \right)$$

otherwise cancel with $\rho = 1$ (coalescence)

Joe's effect on my own trajectory

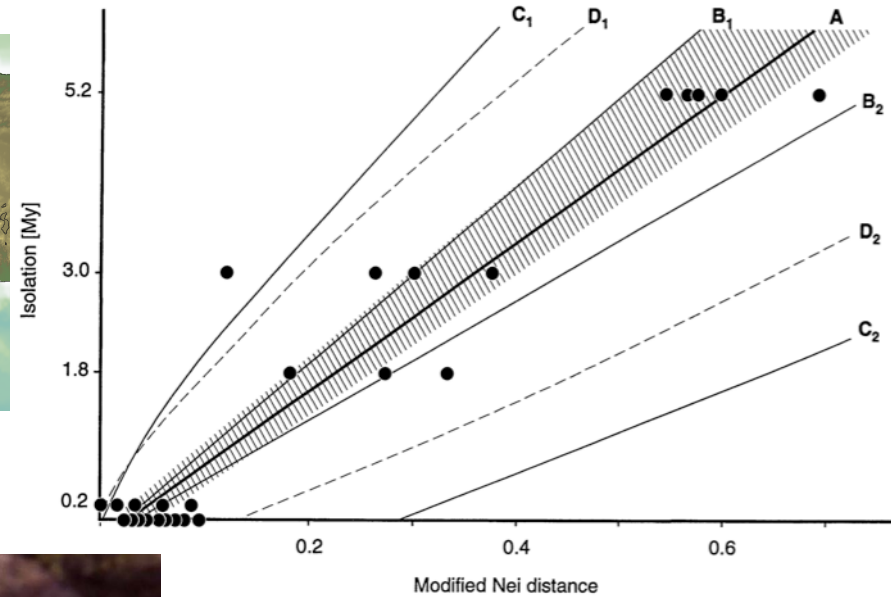
- ◆ Why I wanted to work with Joe
- ◆ While I was in Joe's lab
- ◆ What effect Joe had on me

Multi-locus divergence time estimation 1994



Phylip 3.5c:
contml
using allele frequencies
of 31 allozyme loci

CALIBRATION OF A MOLECULAR CLOCK



GEOLOGICALLY DATED SEA BARRIERS CALIBRATE A PROTEIN CLOCK FOR AEGEAN WATER FROGS

PETER BEERLI,^{1,2} HANSJÜR G HOTZ,^{1,3} AND THOMAS UZZELL⁴

¹Zoologisches Museum, Universität Zürich, Switzerland

³E-mail: hotz@zoolmus.unizh.ch

⁴Department of Ecology, Ethology, and Evolution, University of Illinois, Urbana, Illinois 61821

E-mail: uzzell@uiuc.edu

Working with Joe on the coalescent

and Mary and Jon

We infer parameters Θ from genetic data D by collecting genealogies G

$$G_1, G_2, G_3, \dots, G_i, \dots \sim H(G_{i-1}, \Theta_0)$$

$$L(\Theta) \simeq \frac{1}{n} \sum_i^n p(D|G_i) \frac{f(G_i|\Theta)}{H(G_i)},$$

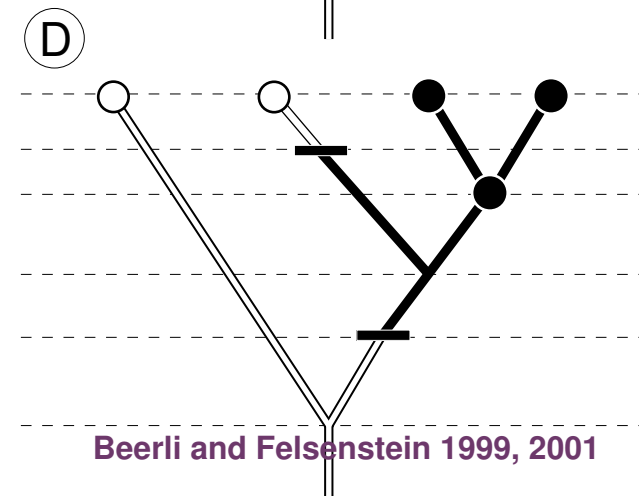
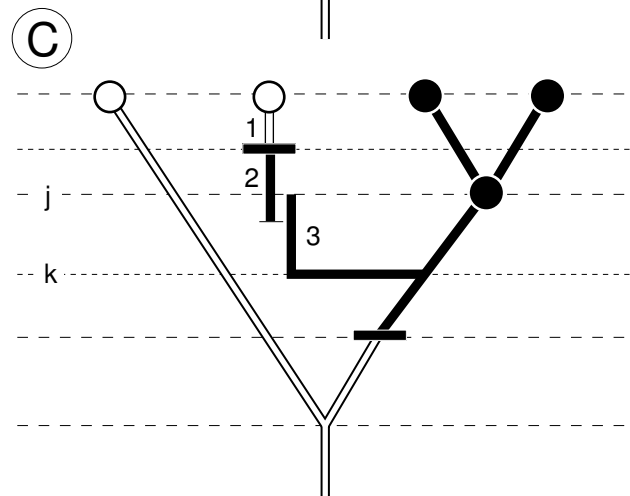
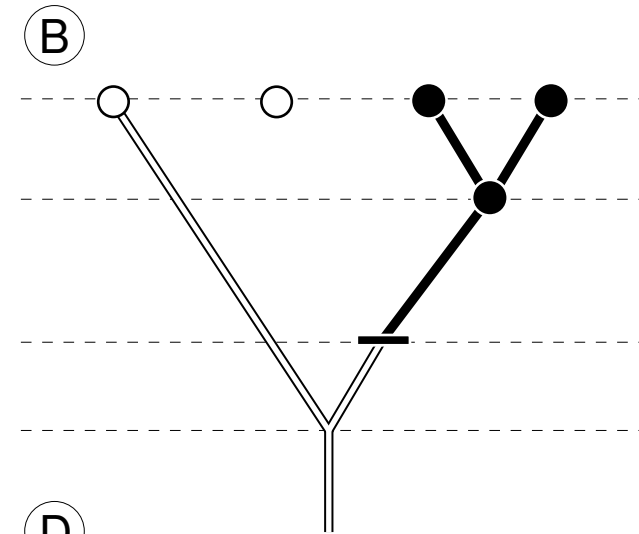
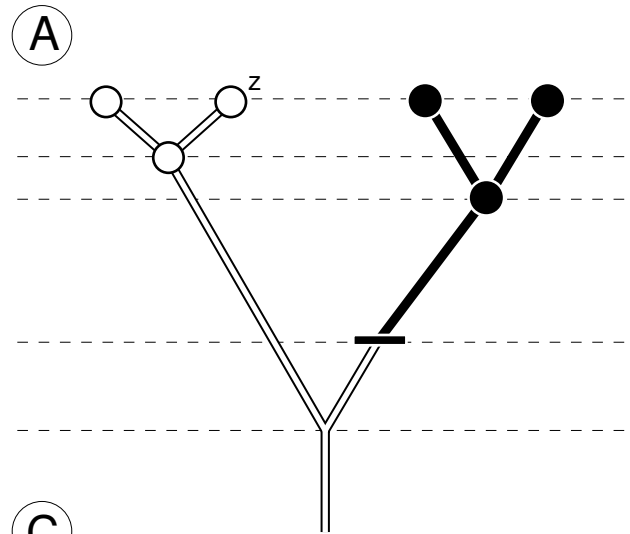
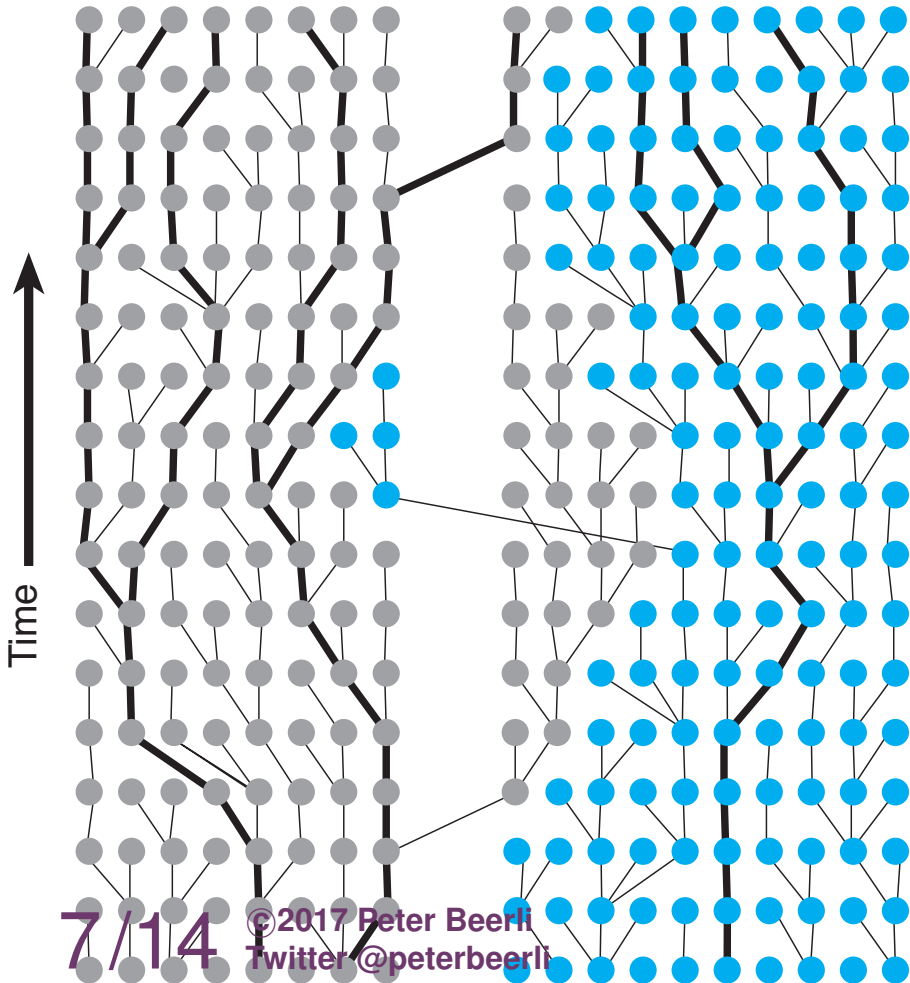
$$\text{a sensible choice is } H \sim \frac{f(G|\Theta_0)p(D|G)}{L(\Theta_0)} \Rightarrow L(\Theta) \simeq \frac{1}{n} \sum_i^n \frac{p(D|G_i)f(G_i|\Theta)L(\Theta_0)}{f(G|\Theta_0)p(D|G_i)}$$

where $f(G|\Theta_0)$ is the probability density of G using the coalescent

$$\hat{\Theta} = \operatorname{argmax}_{\Theta} \frac{L(\Theta)}{L(\Theta_0)} = \operatorname{argmax}_{\Theta} \frac{1}{n} \sum_i^n \frac{f(G_i|\Theta)}{f(G|\Theta_0)}$$

Working with Joe on the coalescent

Structured coalescent



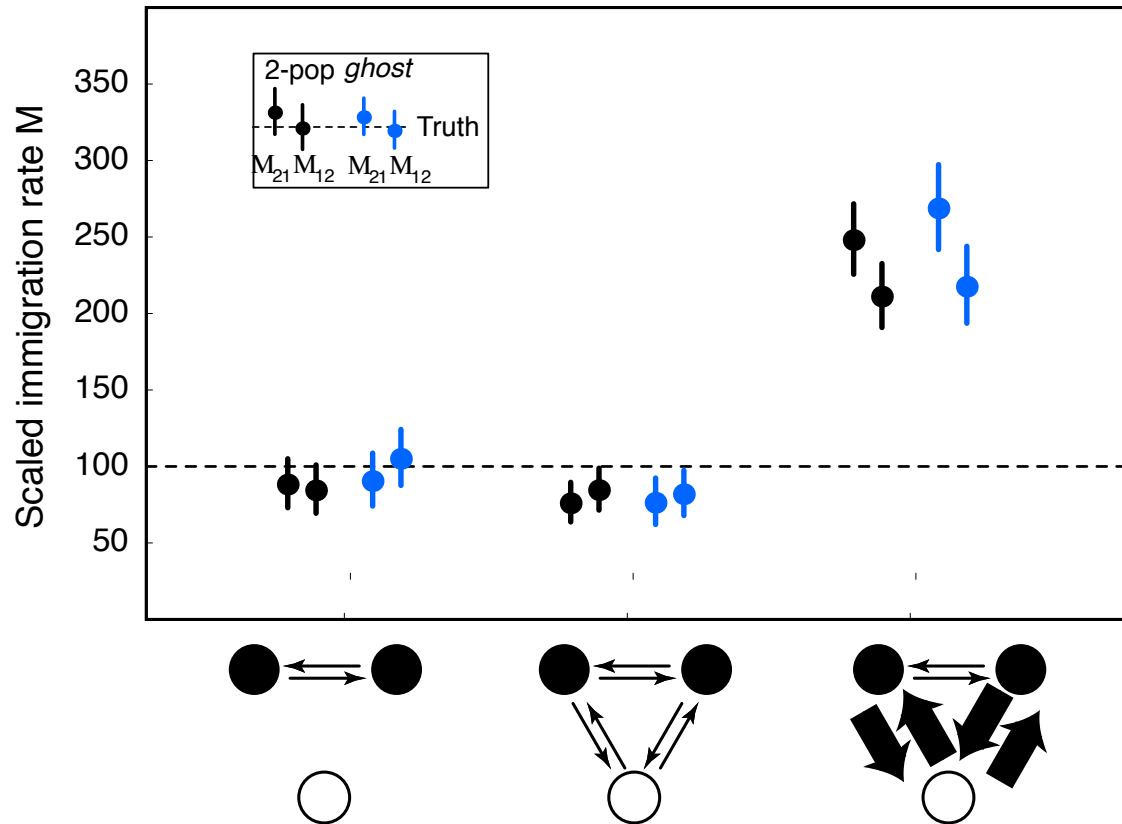
7/14

©2017 Peter Beerli
Twitter @peterbeerli

Beerli and Felsenstein 1999, 2001

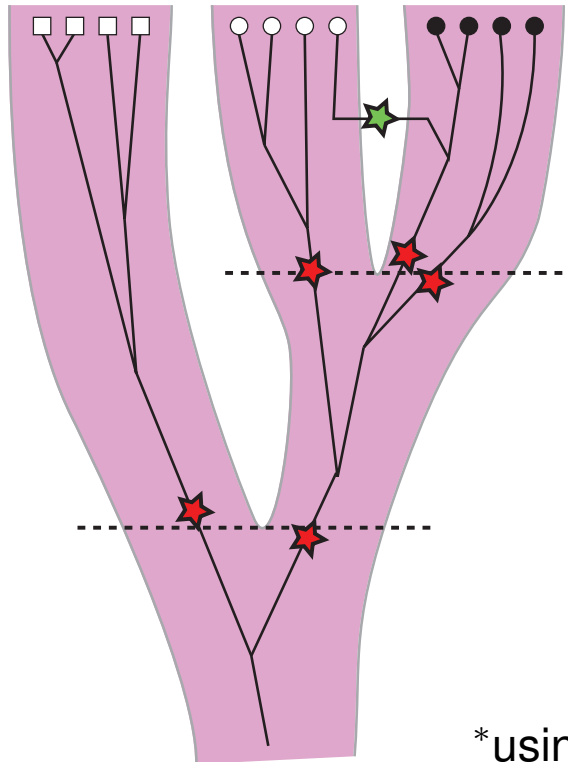
Ghost populations

Structured coalescent



Population splitting

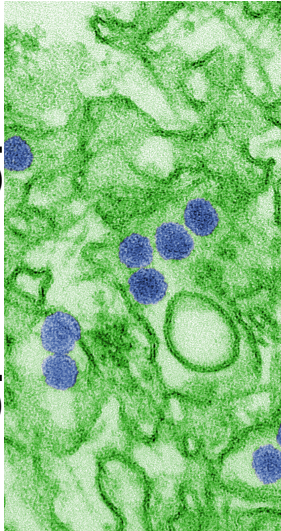
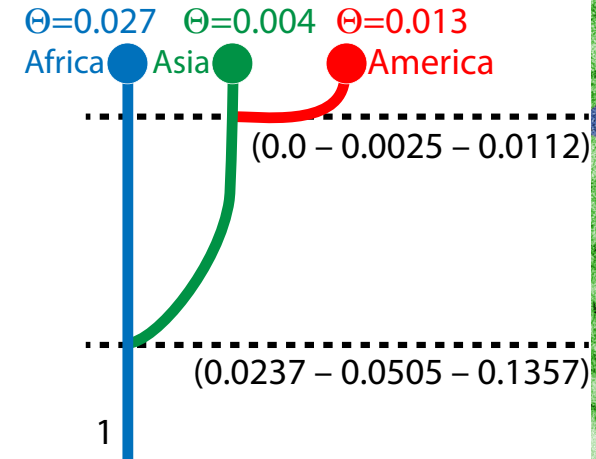
The events in a coalescent tree can be expressed as a waiting time process with rate λ for N populations and k_j lineages in population j :



$$\lambda_{\text{coalescence}} = \sum_{j=1}^N \frac{k_j(k_j - 1)}{4N}$$

$$\lambda_{\text{migration}} = \sum_{j=1}^N \sum_{i=1, i \neq j}^N k_j m_{ij}$$

$$\lambda_{\text{splitting}}^* = \frac{\sqrt{\frac{2}{\pi}} e^{-\frac{(t-\mu)^2}{2b^2}}}{b \left(1 - \text{erf} \left(\frac{t-\mu}{\sqrt{2}b} \right) \right)}$$



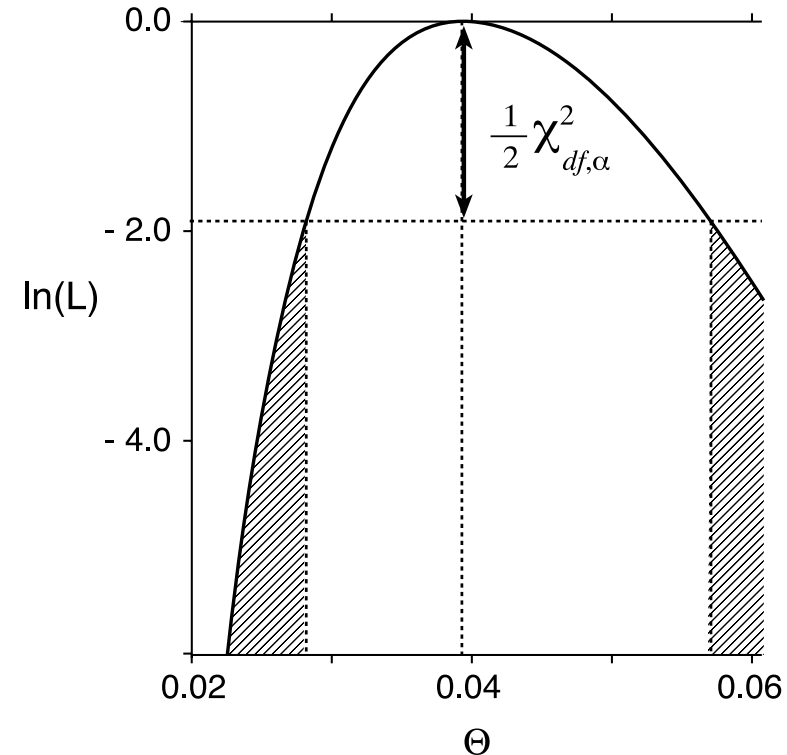
*using a Normal distribution to model the splitting time between two populations.

Likelihood ratio test:

$$H_0 : \Theta_B = (x_1, x_2, 0, \dots) = \Theta_A$$

$$H_1 : \Theta_B = (x_1, x_2, 0, \dots) \neq \Theta_A$$

$$\begin{aligned} \chi_{df}^2 &= -2 \log \left(\frac{L(\Theta_B)L(\Theta_0)}{L(\Theta_A)L(\Theta_0)} \right) \\ &= -2 \log \left(\frac{\sum_G \frac{p(G|\Theta_B)}{p(G|\Theta_0)}}{\sum_G \frac{p(G|\Theta_A)}{p(G|\Theta_0)}} \right) \end{aligned}$$



Population models comparison

Akaike's information criterion:

$$AIC(x) = -2 \log L(\Theta_x) + 2k_x$$

$$\text{Best AIC} = \min_x (AIC(x))$$

Traversing all migration model patterns:

AIC(0++) $\xrightarrow{\text{Yes}}$ AIC(00+) $\xrightarrow{\text{Yes}}$ AIC(000) \rightarrow done

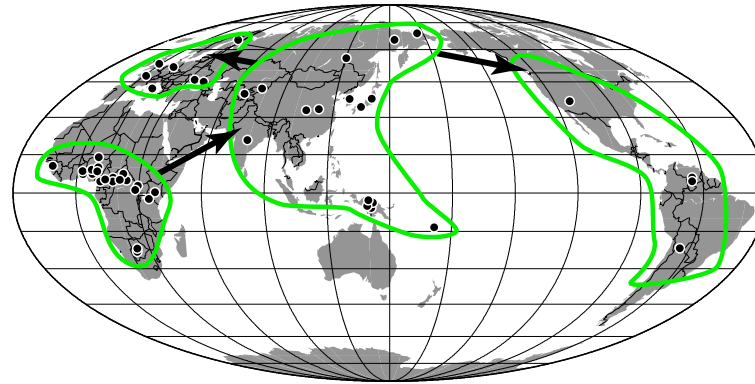
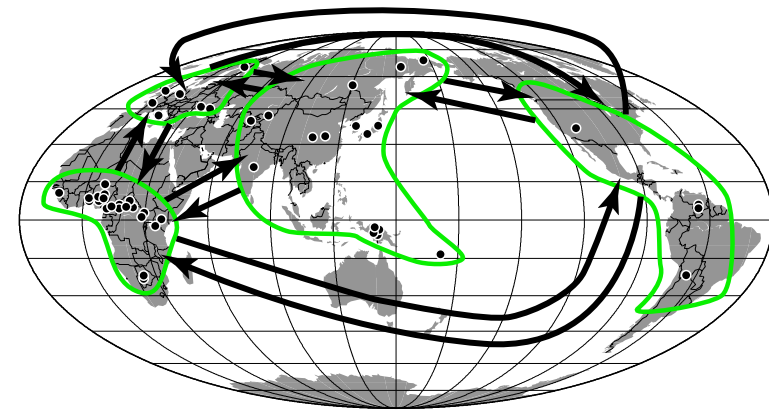
$\xrightarrow{\text{No}}$ done

$\xrightarrow{\text{No}}$ AIC(+0+) $\xrightarrow{\text{Yes}}$ AIC(+00) \rightarrow done

$\xrightarrow{\text{No}}$ AIC(++0) \rightarrow done

[There may be a problem with this approach!]

Full migration model



AFTER analysis, best model based on AIC.

Accept	Model	AIC	#param	Ln L
+	x000 *x** **x* ***x	269.26515	13	-121.6326
+	x000 0x** **x* ***x	267.26515	12	-121.6326
-	x000 0x0* **x* ***x	487.46296	11	-232.7315
+	x000 0x*0 **x* ***x	265.26515	11	-121.6326
-	x000 0x*0 0*x* ***x	340.28333	10	-160.1417
+	x000 0x*0 *0x* ***x	263.26515	10	-121.6326
+	x000 0x*0 *0x0 ***x	261.26515	9	-121.6326
Best	x000 0x*0 *0x0 00*x	257.26515	7	-121.6326

Marginal likelihoods for model selection

Bayesian inference:

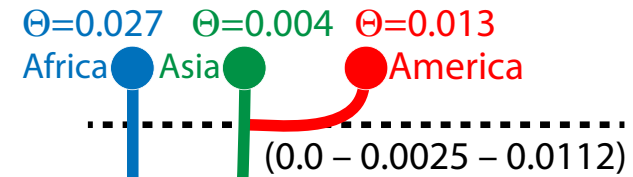
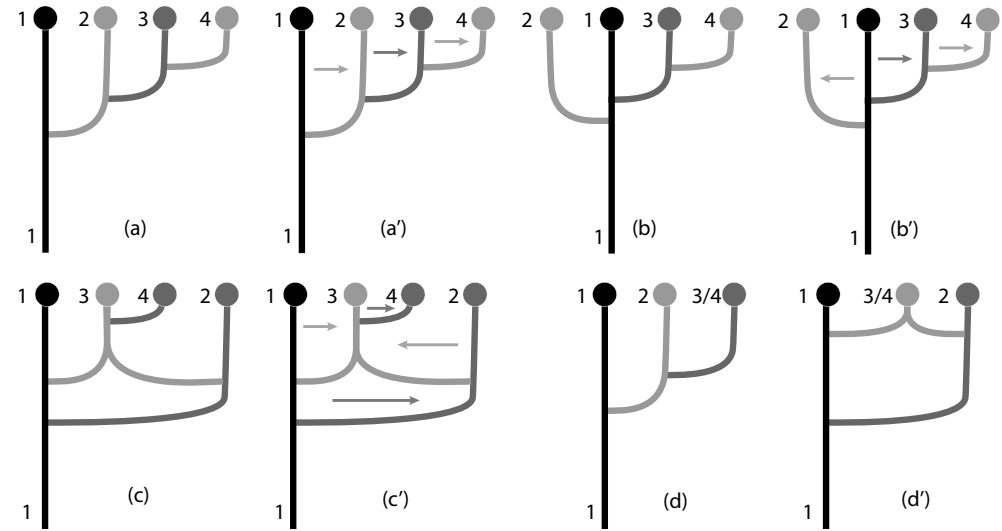
$$p(\Theta|D, M) = \frac{p(\Theta)p(D|\Theta)}{p(D|M)} = \frac{p(\Theta)p(D|\Theta)}{\int_{\Theta} p(\Theta)p(D|\Theta)d\Theta}$$

Theorem: The combined marginal likelihoods over all independent data blocks can be calculated as a product of independently calculated marginal likelihoods for each data block and a constant. [Proof: Beerli and Palczewski 2010]

$$P(D_1, \dots, D_n|M) = K \prod_{i=1}^n P(D_i|M)$$

$$K = \int_{\theta} \prod_i P(\theta|D_i, M) P(\theta|M)^{1-n} d\theta.$$

This allows the calculation of the combined marginal likelihood using independent inferences.



Some things I wanted to know about Population genetics

$$P(S|G, E) = \frac{\prod_j \frac{k_j(k_j-1)}{\theta_j} \rho_j}{\prod_j \frac{k_j(k_j-1)}{\theta_j} \rho_j + \sum_i \frac{\sum_j k_i M_{ji} (1-\rho_j)}{\theta_i}} + \frac{\sum_i \frac{\sum_j k_i M_{ji} (1-\rho_j)}{\theta_i}}{\prod_j \frac{k_j(k_j-1)}{\theta_j} \rho_j + \sum_i \frac{\sum_j k_i M_{ji} (1-\rho_j)}{\theta_i}}$$

so we can assemble all parts to get the final result

there are $\frac{2}{k_i(k_i-1)}$ per coalescent events on interval and M_{ji} per lineage interval

$$\exp\left(-u \left(\sum_i \frac{k_i(k_i-1)}{\theta_i} + \sum_i \sum_j k_i M_{ji} \right) \left(\sum_i \frac{k_i(k_i-1)}{\theta_i} + \sum_i \sum_j k_i M_{ji} \right) \right) * \begin{cases} 1 & \text{if coalescence} \\ \rho & \text{if coalescence, otherwise} \end{cases}$$

$$\left(\frac{\prod_i \frac{k_i(k_i-1)}{\theta_i} \rho_i}{\sum_i \frac{k_i(k_i-1)}{\theta_i} + \sum_i \sum_j k_i M_{ji}} + \frac{\sum_i \sum_j k_i M_{ji} (1-\rho_j)}{\sum_i \frac{k_i(k_i-1)}{\theta_i} + \sum_i \sum_j k_i M_{ji}} \right) * \left(\frac{\frac{k_j(k_j-1) \rho_j}{\theta_j} \frac{2}{k_j(k_j-1)}}{\sum_i \frac{k_i(k_i-1)}{\theta_i} + \sum_i \sum_j k_i M_{ji}} + \frac{\sum_i \sum_j k_i M_{ji} (1-\rho_j)}{\sum_i \frac{k_i(k_i-1)}{\theta_i} + \sum_i \sum_j k_i M_{ji}} \right)$$

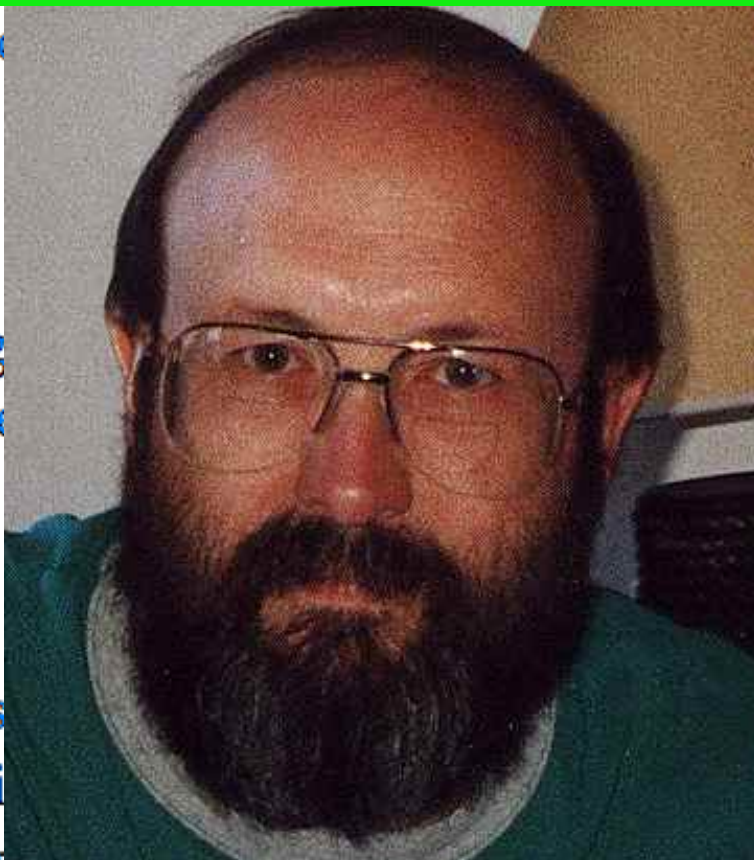
otherwise cancel with $\rho = 1$ (coalescence)

I learned a lot about population genetics – THANK YOU JOE!!

coalesc

Change
change
with

The es
quantit



$$L(D) = \int_0^\infty \frac{1}{2N} e^{-\frac{t}{2N}} \text{Prob}(D | \mu t, 1) dt.$$

scale u which is measured in expected nu
replace t by u/μ and dt by $(1/\mu) du$. This

$$L(D) = \int_0^\infty \frac{1}{2N\mu} e^{-\frac{u}{2N\mu}} \text{Prob}(D | u, 1) du.$$

at the likelihood turns out to be a function,
only of their product $N\mu$.

distributions of divergence time ut for three