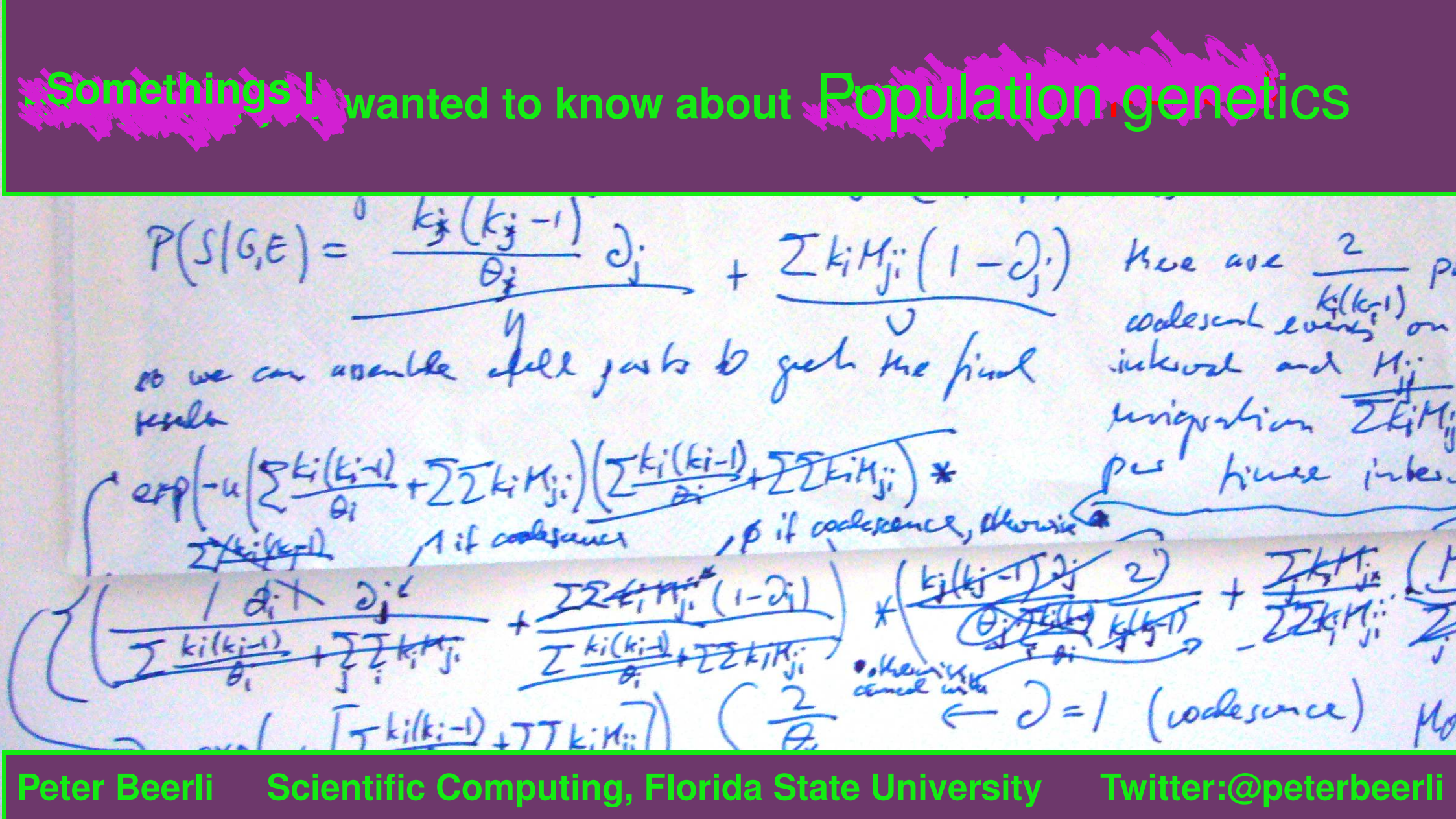
Everything you wanted to know about Frogs

Peter Beerli Scientific Computing, Florida State University Twitter:@peterbeerli

Everything you wanted to know about Computers

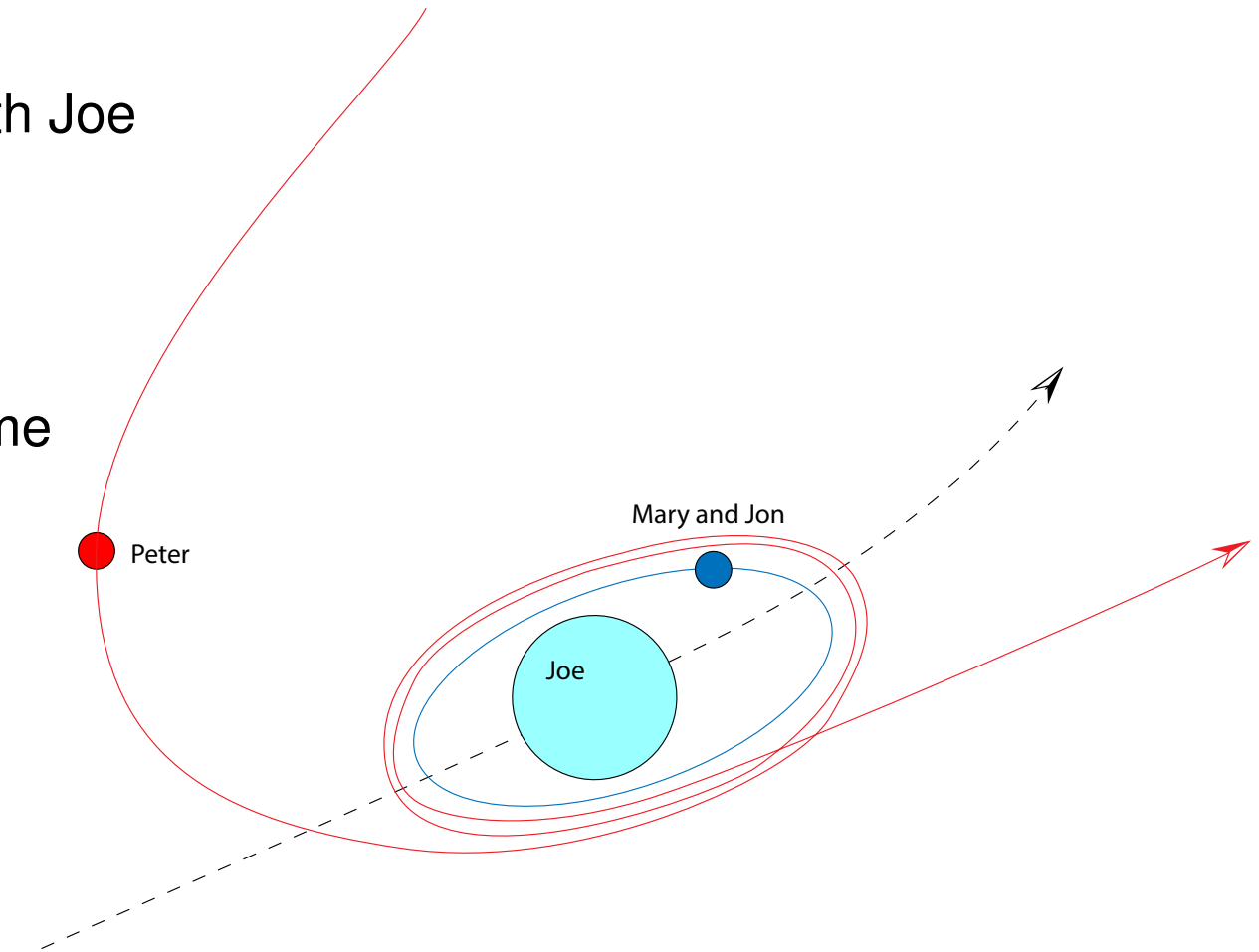
Peter Beerli Scientific Computing, Florida State University Twitter:@peterbeerli

Somethings I wanted to know about Population genetics



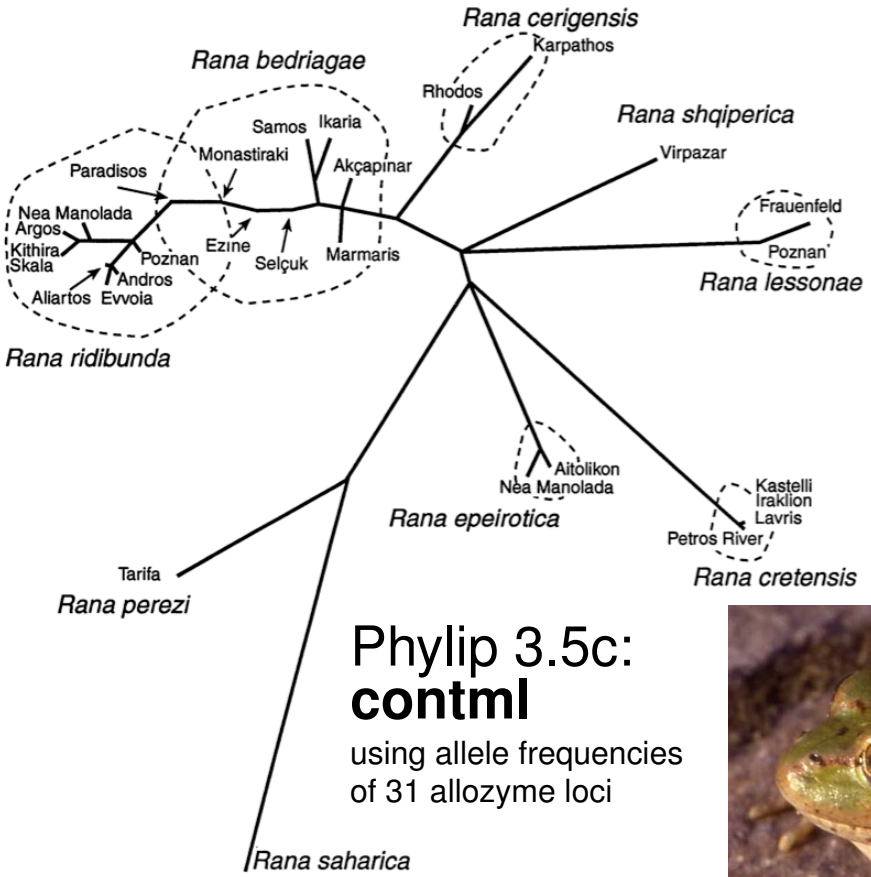
Joe's effect on my own trajectory

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

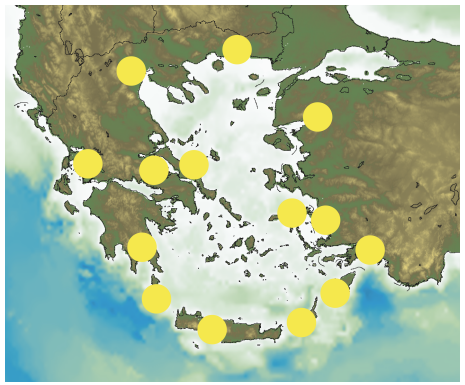


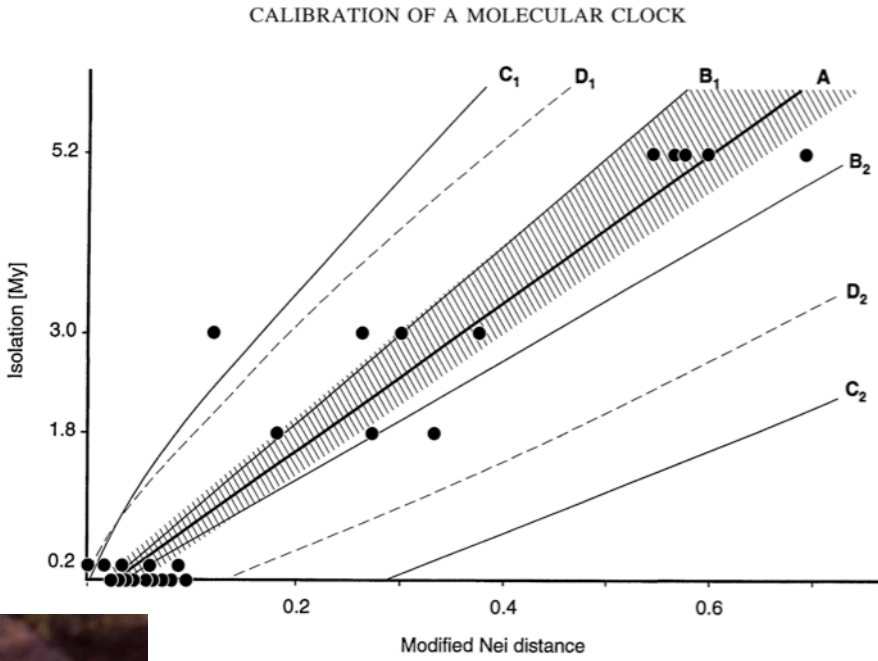
©2017 Peter Beerli
Twitter @peterbeerli

Multi-locus divergence time estimation 1994



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





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

PETER BEERLI,[1,2] HANSJÜRG HOTZ,[1,3] AND THOMAS UZZELL[4]
[1]Zoologisches Museum, Universität Zürich, Switzerland
[3]E-mail: hotz@zoolmus.unizh.ch
[4]Department of Ecology, Ethology, and Evolution, University of Illinois, Urbana, Illinois 61821
E-mail: uzzell@uiuc.edu

Beerli, Hotz, and Uzzell 1996

©2017 Peter Beerli
Twitter @peterbeerli

Working with Joe on the coalescent

and Mary and Jon

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

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

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

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

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

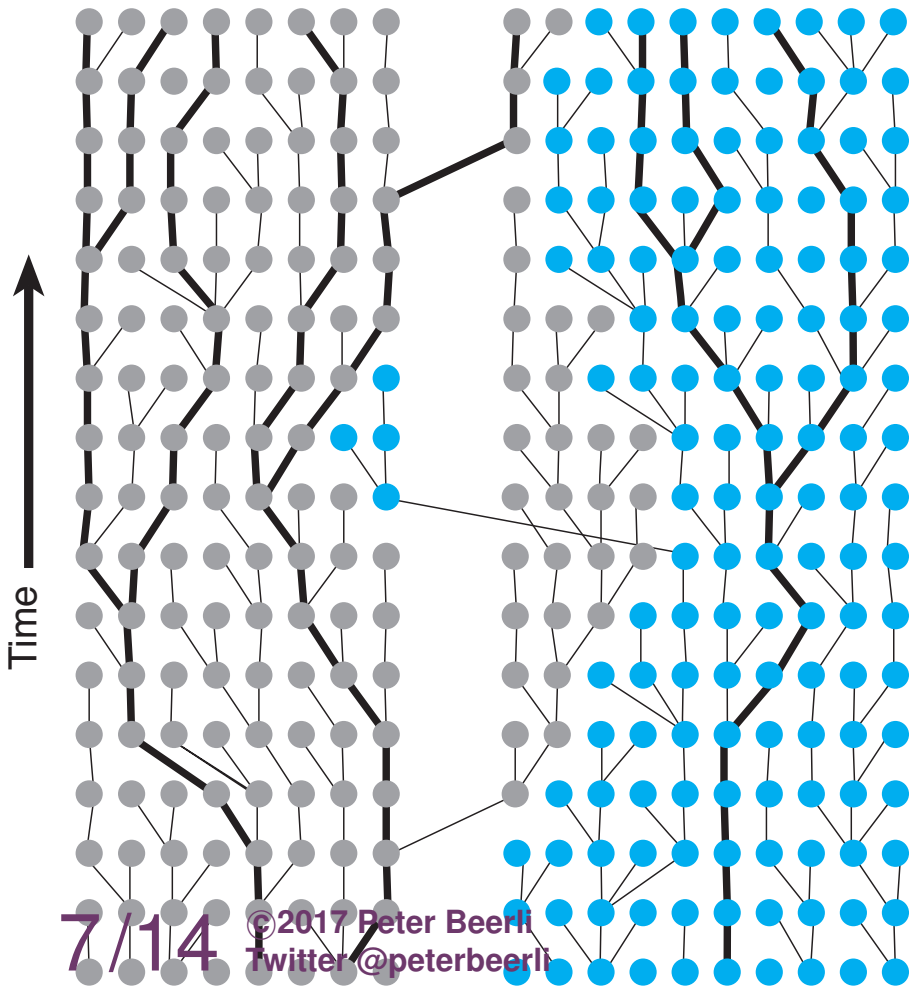
$$\hat{\Theta} = \underset{\Theta}{\mathrm{argmax}}\,\frac{L(\Theta)}{L(\Theta_0)} = \underset{\Theta}{\mathrm{argmax}}\,\frac{1}{n}\sum_i^n \frac{f(G_i|\Theta)}{f(G|\Theta_0)}$$

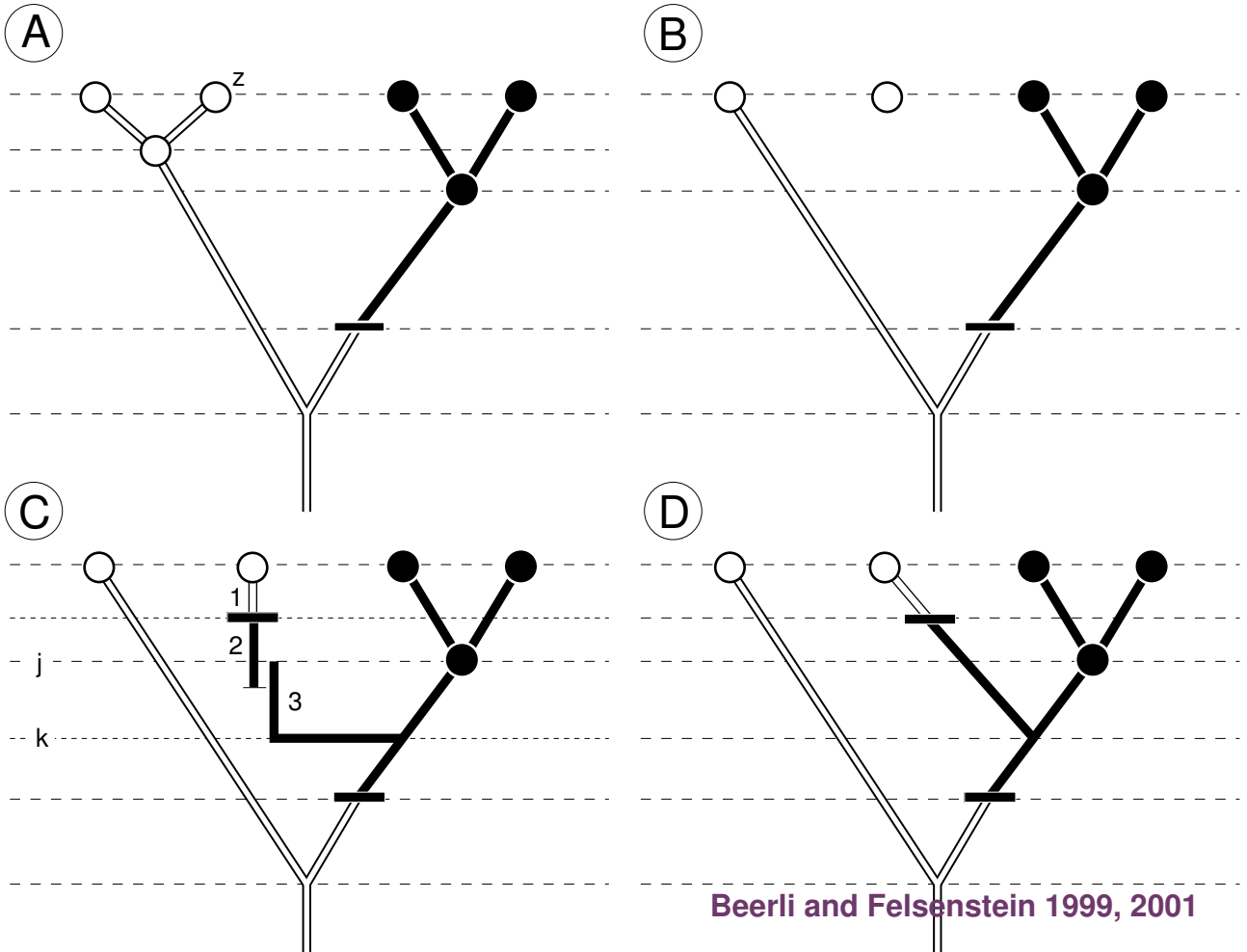
©2017 Peter Beerli
Twitter @peterbeerli

Felsenstein, Kuhner, Yamato, and Beerli 1999

Working with Joe on the coalescent

Structured coalescent

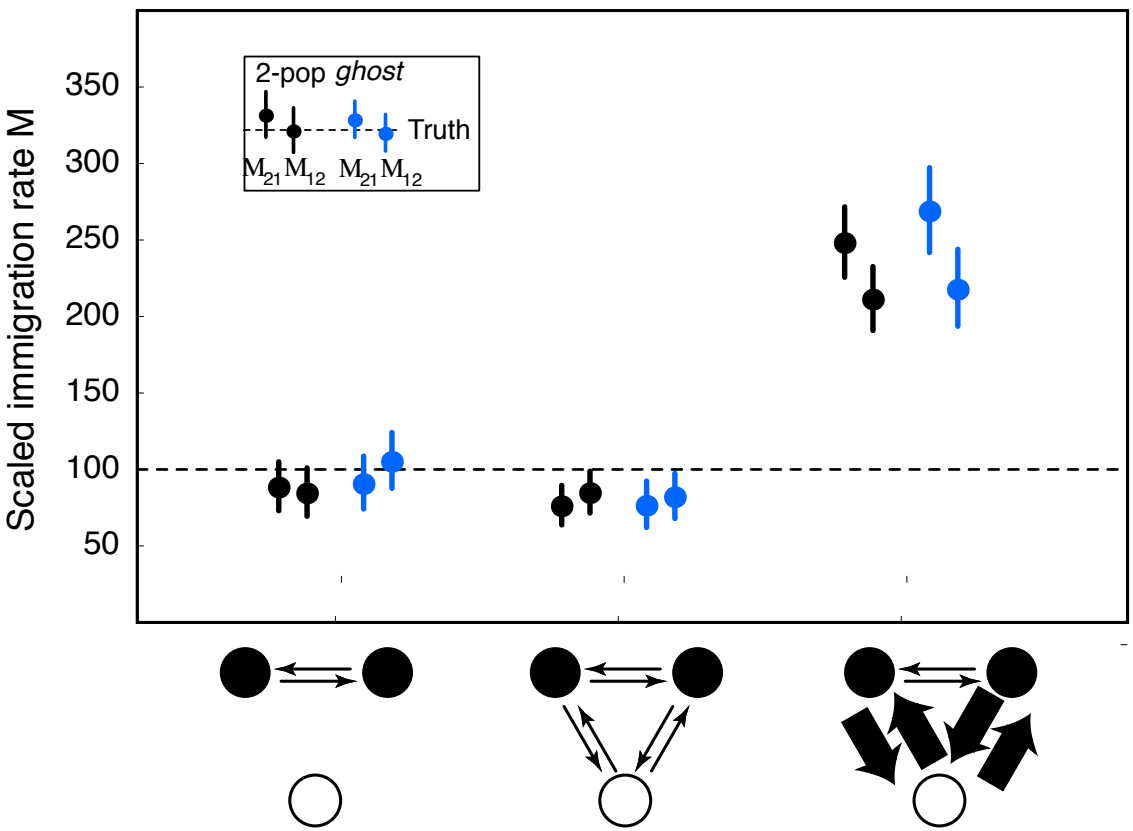




Beerli and Felsenstein 1999, 2001

©2017 Peter Beerli
Twitter @peterbeerli

Ghost populations

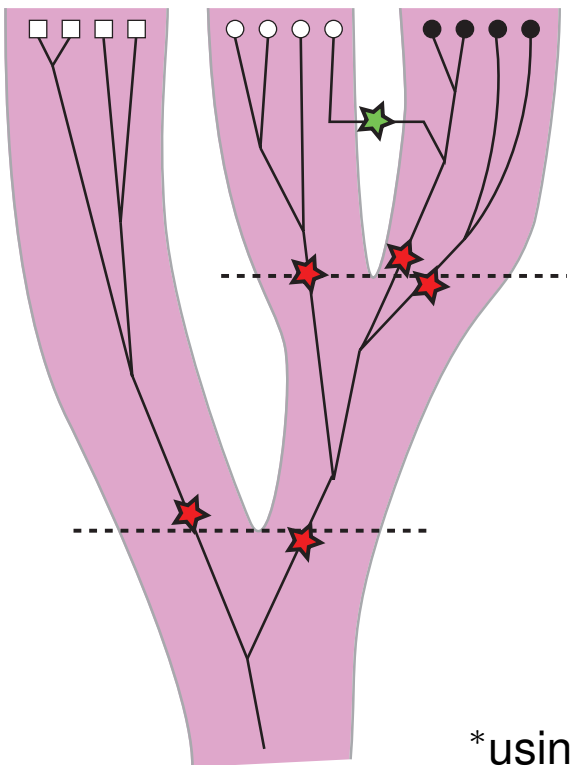


©2017 Peter Beerli
Twitter @peterbeerli

Beerli 2004

Population splitting

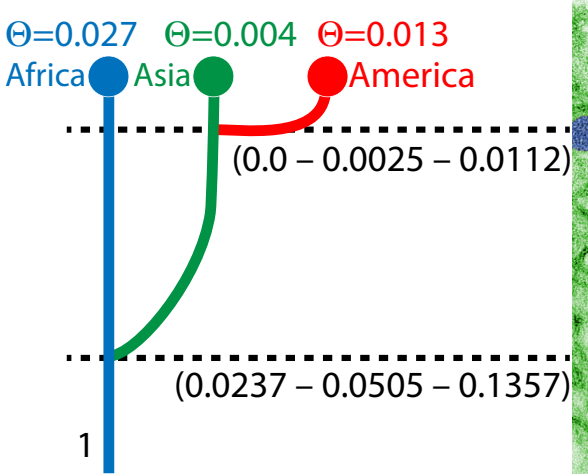
The events in a coalescent tree can be expressed as a waiting time process with rate λ for N populations and k_j lineages in population j:

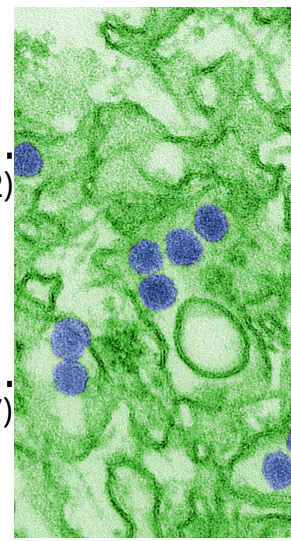


$$\lambda_{\text{coalescence}} = \sum_{j=1}^{N} \frac{k_j(k_j-1)}{4N}$$

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

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





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

©2017 Peter Beerli
Twitter @peterbeerli

Population models comparison

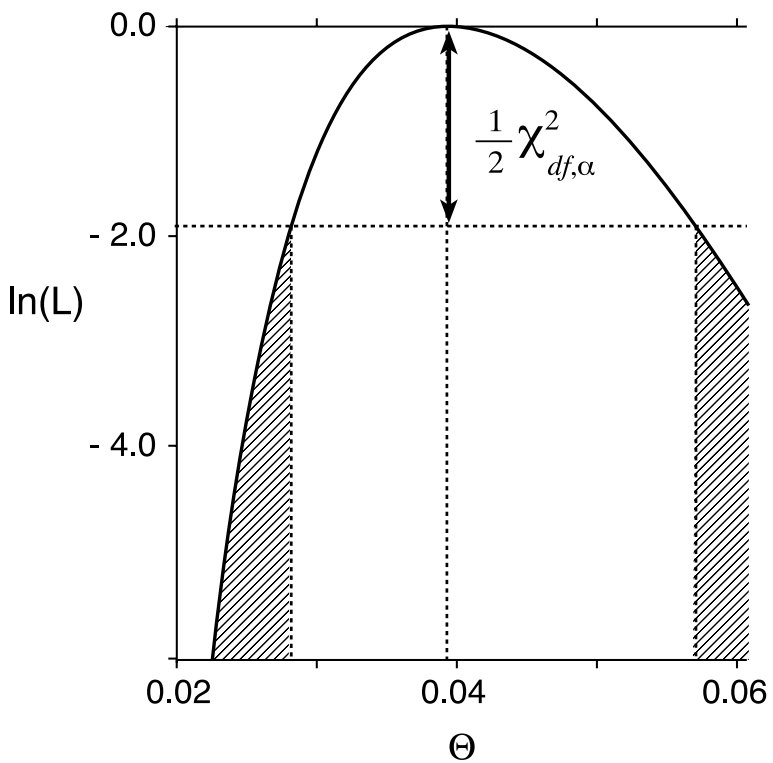
Likelihood ratio test:

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

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

$$\chi^2_{df} = -2 \log \left(\frac{\mathrm{L}(\Theta_B)\mathrm{L}(\Theta_0)}{\mathrm{L}(\Theta_A)\mathrm{L}(\Theta_0)} \right)$$

$$= -2 \log \left(\frac{\sum_G \frac{\mathrm{p(G|\Theta_B)}}{\mathrm{p(G|\Theta_0)}}}{\sum_G \frac{\mathrm{p(G|\Theta_A)}}{\mathrm{p(G|\Theta_0)}}} \right)$$



©2017 Peter Beerli
Twitter @peterbeerli

Population models comparison

Akaike's information criterion:

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

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

Traversing all migration model patterns:

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

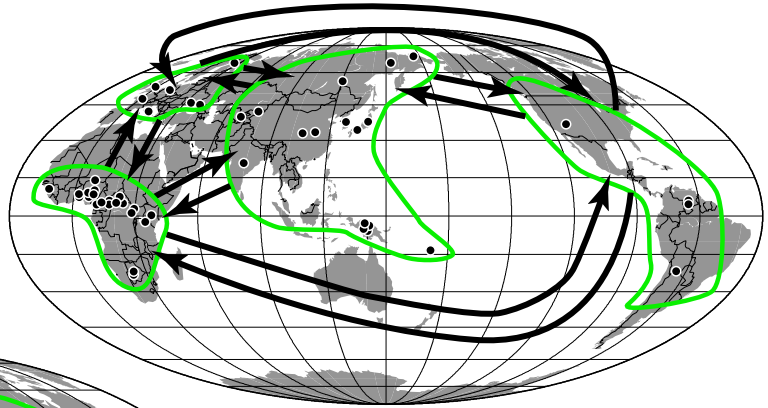
AIC(00+) $\xrightarrow{\text{No}}$ done

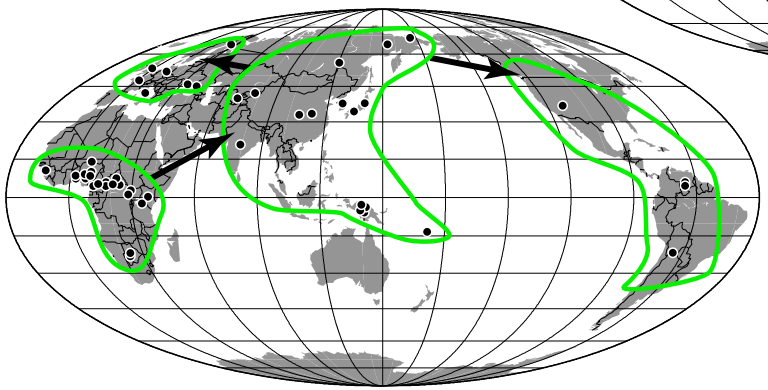
AIC(0++) $\xrightarrow{\text{No}}$ AIC(+0+) $\xrightarrow{\text{Yes}}$ AIC(+00) $\longrightarrow$ done

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

in MIGRATE in 2001

©2017 Peter Beerli
Twitter @peterbeerli

Marginal likelihoods for model selection

Bayesian inference:

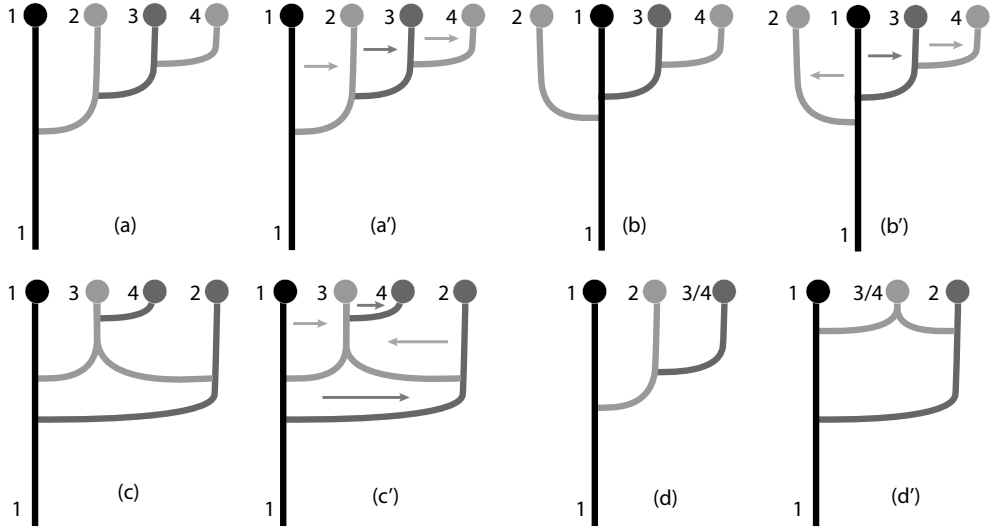
$$\mathrm{p}(\Theta|\mathrm{D},\mathrm{M}) = \frac{\mathrm{p}(\Theta)\mathrm{p}(\mathrm{D}|\Theta)}{\mathrm{p}(\mathrm{D}|\mathrm{M})} = \frac{\mathrm{p}(\Theta)\mathrm{p}(\mathrm{D}|\Theta)}{\int_\Theta \mathrm{p}(\Theta)\mathrm{p}(\mathrm{D}|\Theta)\mathrm{d}\Theta}$$

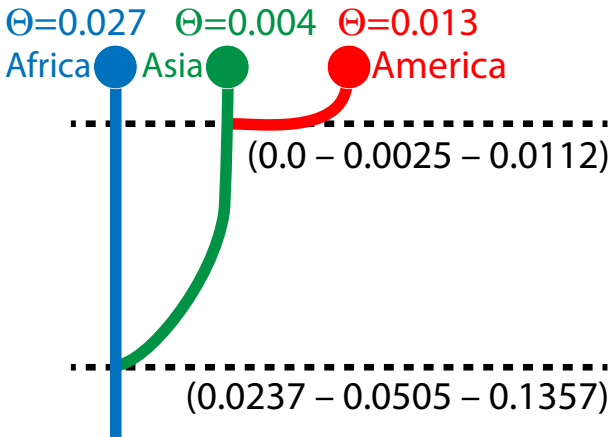
The combined marginal likelihoods over all independent data blocks can be calculated as a product of independently calculated marginal likelihoods for each data block and a constant.

$$P(D_1, \ldots, D_n|M) = K \prod_i^n P(D_i|M)$$

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

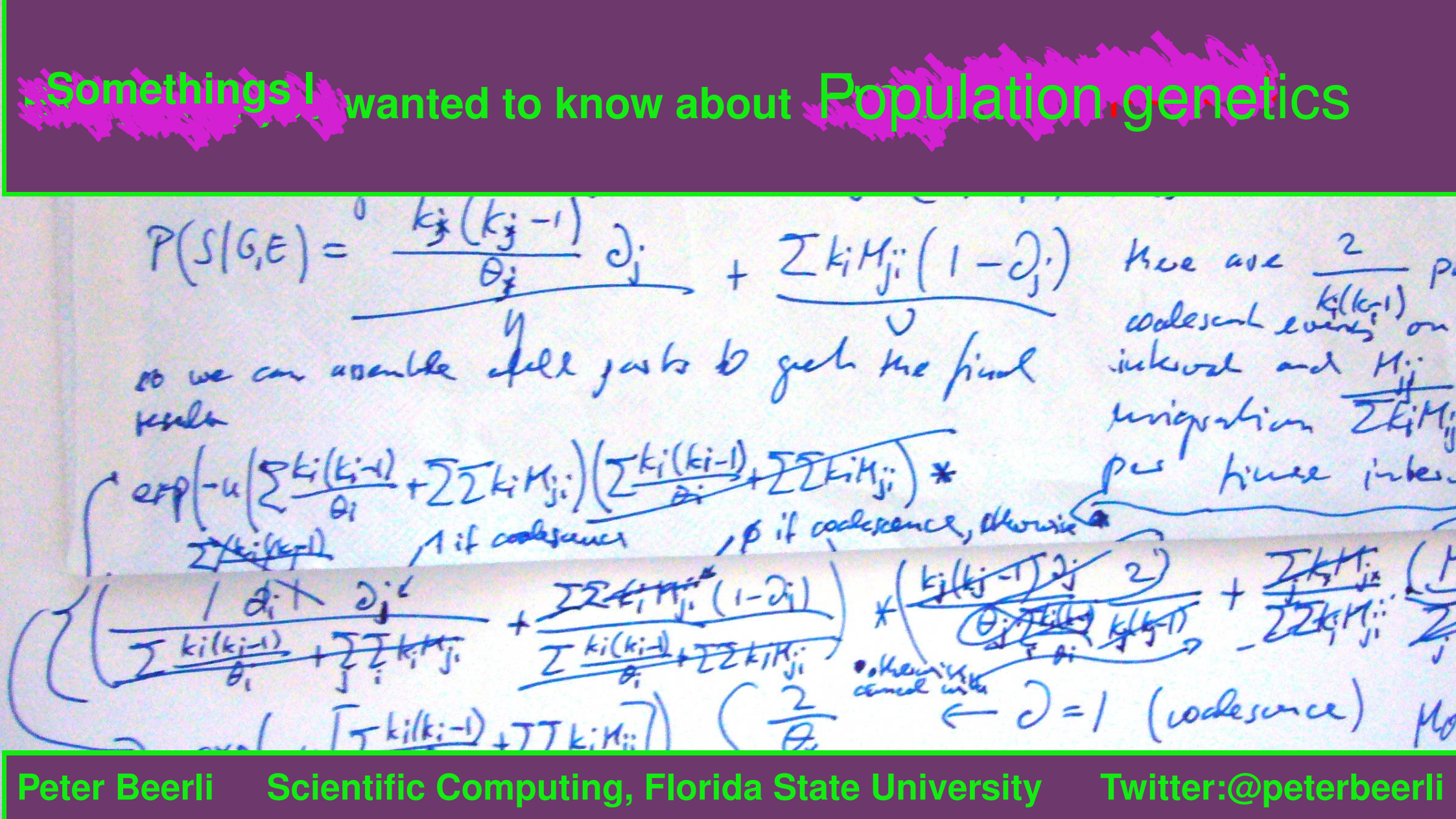
This allows the calculation of the combined marginal likelihood using independent inferences.





©2017 Peter Beerli
Twitter @peterbeerli
Beerli and Palczewski 2010

Somethings I wanted to know about Population genetics



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

and also thanks to Mary Kuhner and Jon Yamato

and SNF, NFS, NIH, UW, FSU