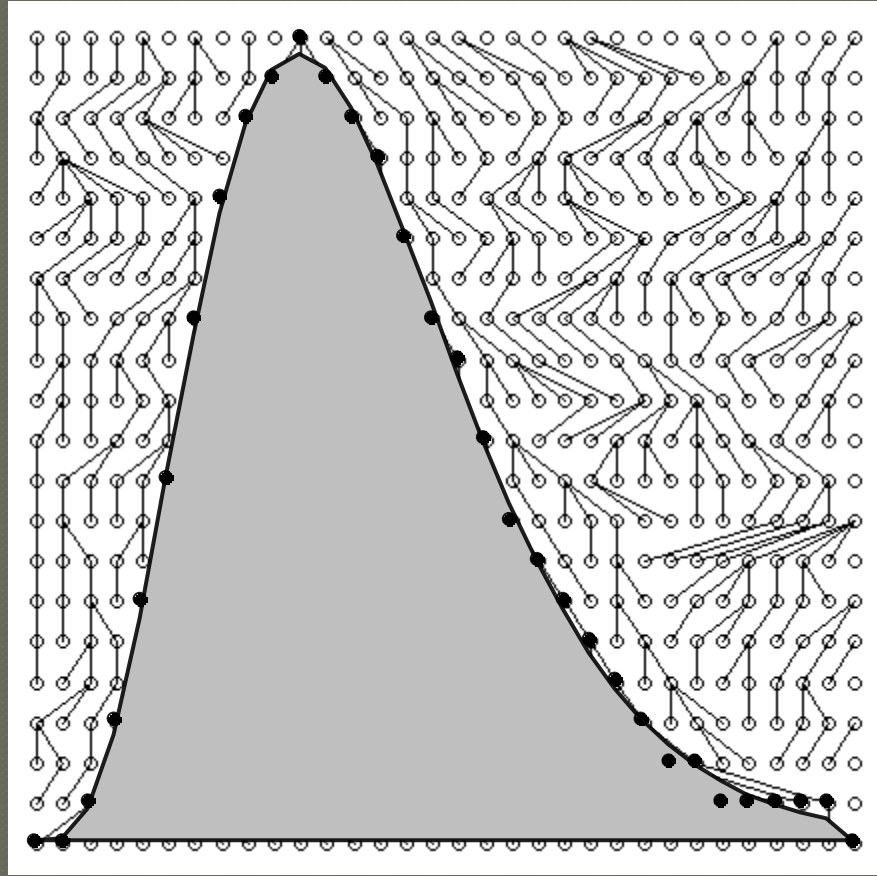


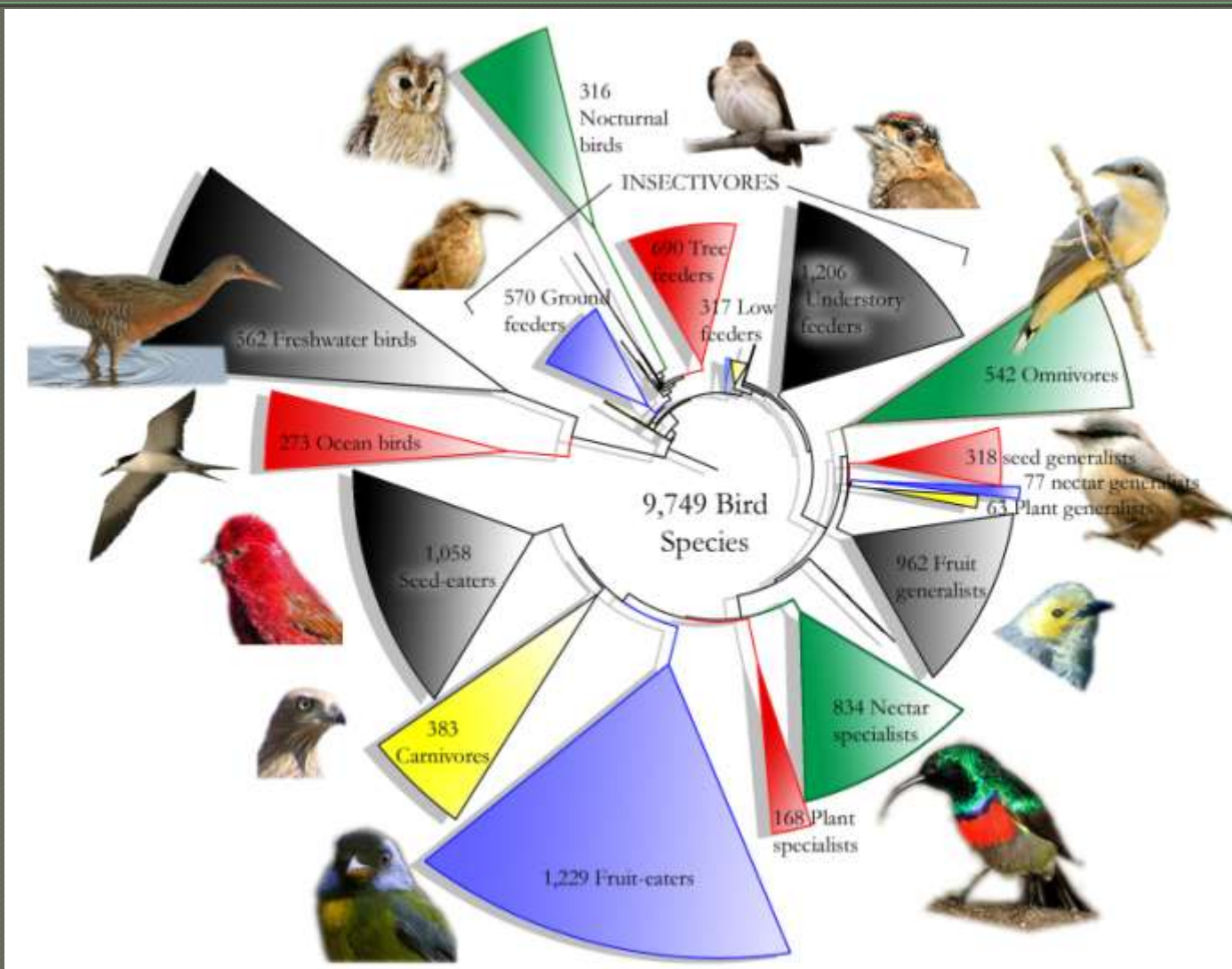


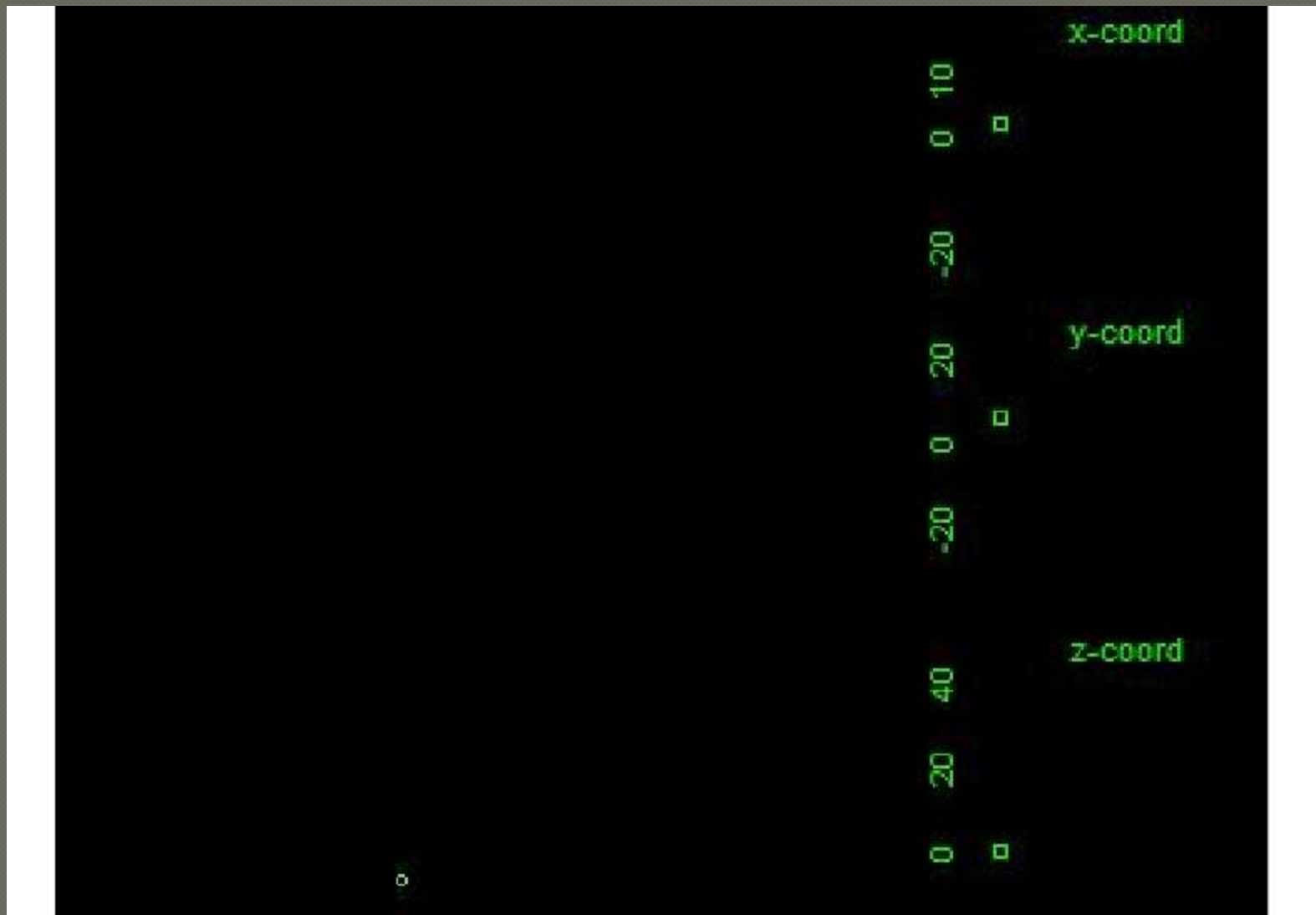
Complex scenarios and
Bayesian Coalescence

παλεστίν κοαλεσσενς



What I normally do





Phylochronology

- Reconstruct the history of related individuals
- Use a comparative hypothesis framework

Site 1



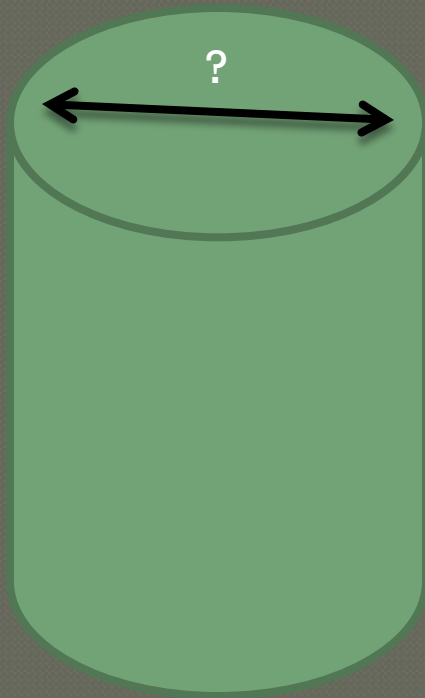
High genetic diversity
(7 haplotypes)

Site 2

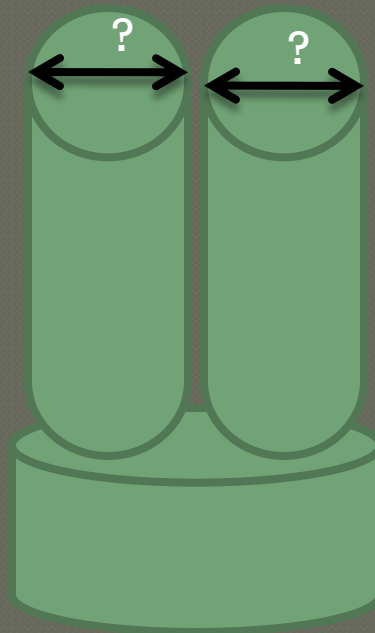


Low genetic diversity
(3 haplotypes)

Alternative Hypotheses



Hypothesis 1:
One large
population



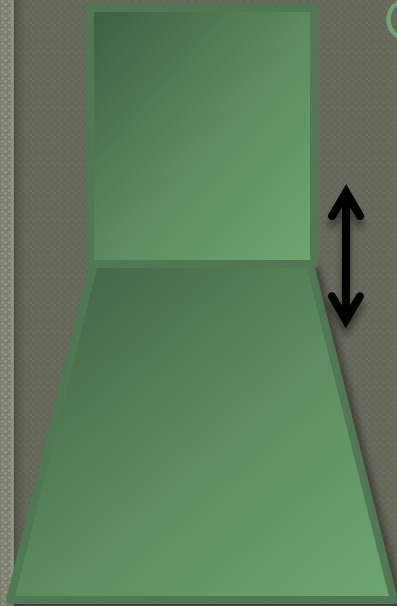
Hypothesis 2:
Two long-
separated
populations

Goals

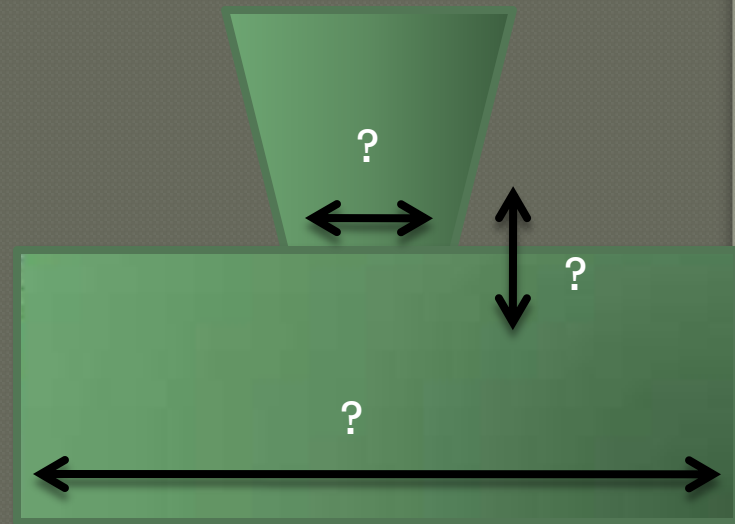
- 1) Find best parameters of these models
- 2) Compare the likelihoods of the two models

For example...

present

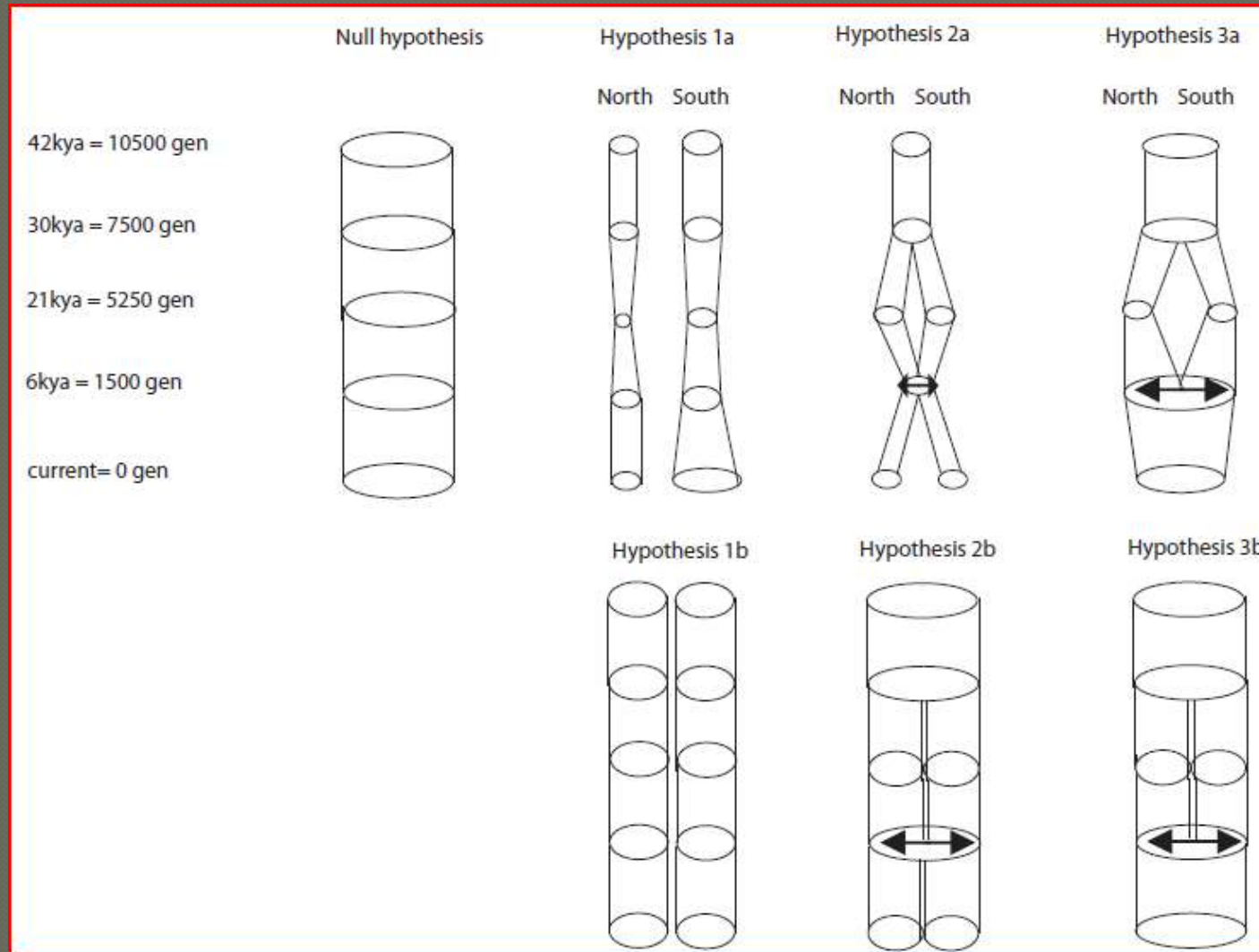


- When did bison start to decline? (~Shapiro *et al*, 2004)

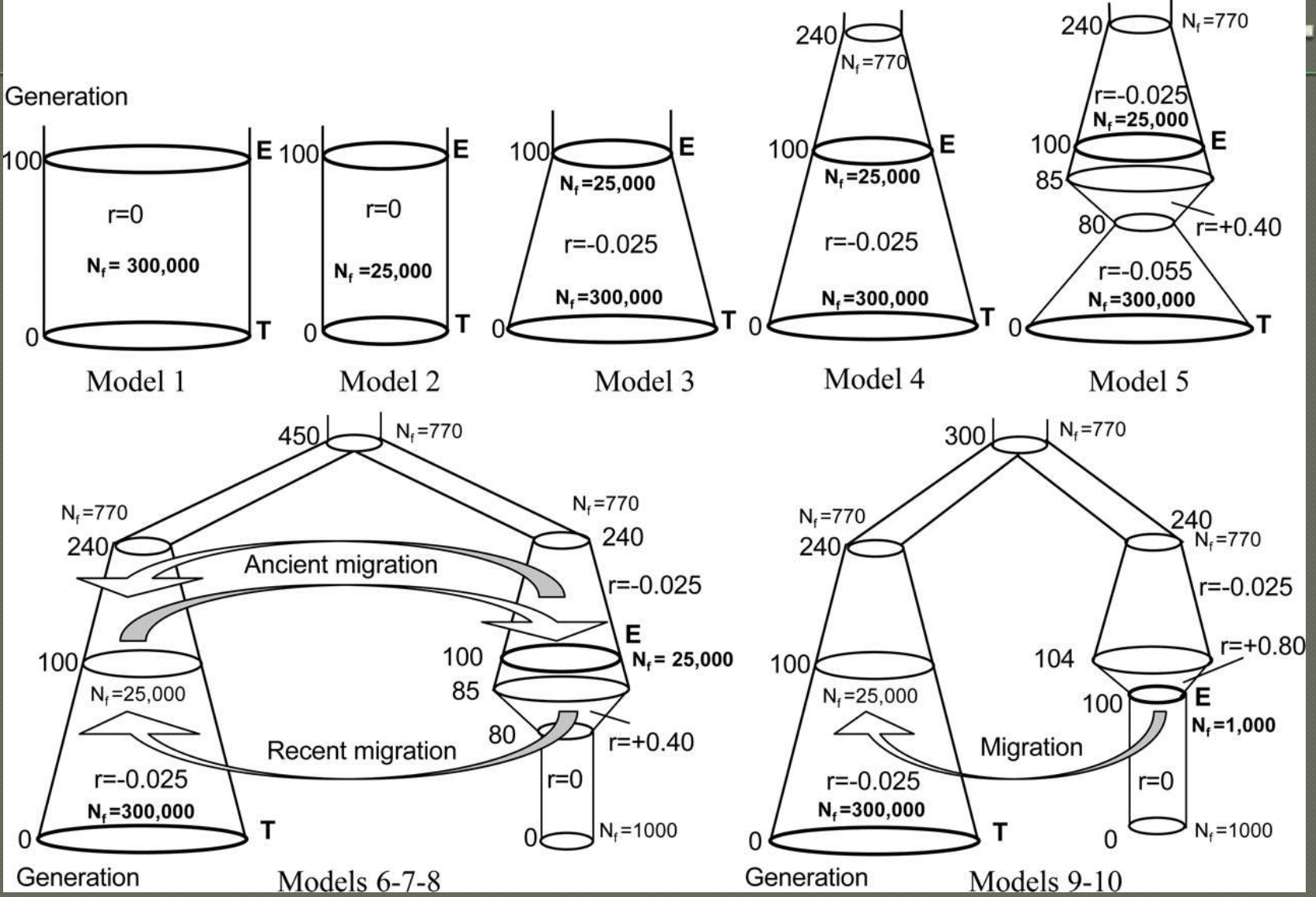


- Did tuco-tucos go through a bottleneck? (Chan *et al*, 2006)

Examples from Jessica



For example



Doodle time!

- Take a few minutes to draw cartoons of alternative models of your data
- Things to keep in mind:
 - Population structure
 - Growth rates
 - Migration
 - Time of events
 - Severity of events
 - Multiple changes through time

Simulation and Simulacrum

We start with a
population that exists
now (generation 0)



Parameter

Symbol

Time

t

0

Scenario testing

SITE 1



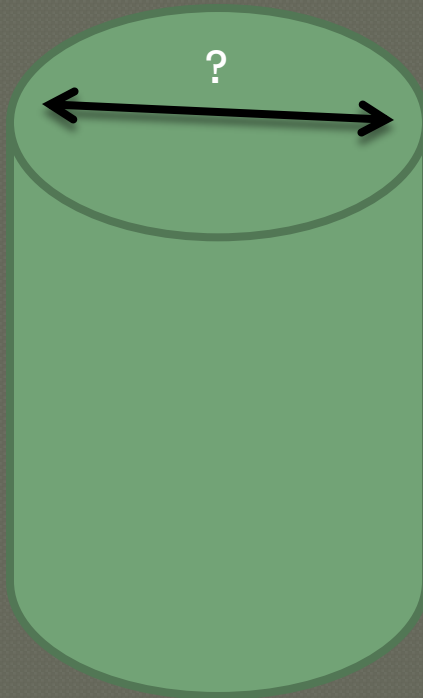
High genetic
diversity

SITE 2

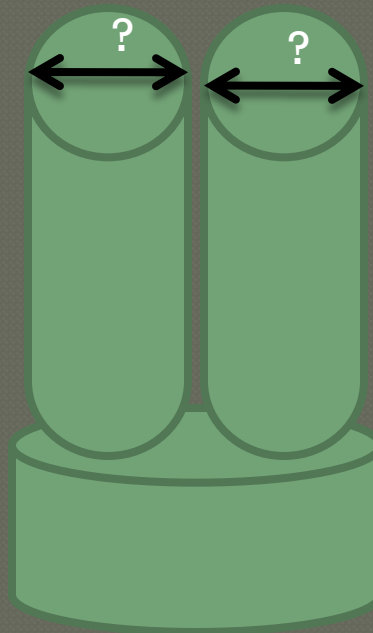


Low
genetic
diversity

Alternative Hypotheses



Hypothesis 1:
One large
population



Hypothesis 2:
Two long-
separated
populations

- 0) Propose a full model, and test for consistency with data
- 1) Find best parameters of these models
- 2) Compare the likelihoods of the two models

Scenario testing (not Bayesian)

Model

.par file, possibly
with free parameters



Scenario testing

H0: The two sites are
one large population
N=1000

H1: The two sites are
two small populations
N=500

H2: A large and a small
population N=750,250

A word about N and Ne

- The census size is rarely very close to the “effective population size”
- The relationship $N=10 \times N_e$ is often seen in the literature, but
 - Ground squirrels
 - Humans
- Can be affected by changes in census size (good), but also migration, mutation rate, and structure

Scenario testing (not Bayesian)

Model

.par file, possibly
with free parameters



Simulated
Results

_stat.csv file

Open .csv file

Columns of summary statistics

GROUP 0	Haptypes	PrivHaps	SegSites	PairDiffs	HapDiver	NucltdDiv	TajimasD	MismatDist0 VS 1	PrivTo0	PrivTo1	PairDiffs	MeanDiv(fPoolDiv(Ht)	Fst
	1	1	0	0	0	0	nan	{15 }	0	0	0	0	nan
	6	6	431	192.267	0.833333	0.312629	0.12092{0	0 0 0 0 0 0 1 0 0 0	4	37	209.769	0.903847	0.94765 0.0069558
	6	6	293	119.067	0.833333	0.193604	-0.468846{0	0 0 0 0 0 0 0 0 0 0	5	37	183.97	0.903189	0.949458 0.0819505
	6	6	77	38.7333	0.833333	0.062981	0.957837{0	1 1 0 0 0 0 0 0 1 0	3	26	39.765	0.894642	0.932364 -0.063578
	6	6	204	79.6667	0.833333	0.129539	-0.703161{0	0 0 0 0 0 0 0 0 0 1	3	31	78.1368	0.900559	0.937459 0.001258
	6	6	294	121.867	0.833333	0.198157	-0.347997{0	0 0 0 0 0 0 0 0 0 0	4	33	115.295	0.900559	0.946006 -0.05131
	6	6	439	243.867	0.833333	0.396531	1.74664{0	0 0 0 0 1 0 0 0 0 0	6	37	226.479	0.902532	0.951266 -0.050053
	6	6	187	90.7333	0.833333	0.147534	0.700151{0	0 0 0 0 0 0 0 0 0 0	4	34	82.1282	0.901874	0.946663 -0.071951
	6	6	125	59.8	0.833333	0.097236	0.59791{0	0 0 0 0 0 0 0 0 0 0	5	30	52.5385	0.897929	0.944691 -0.035445
	6	6	337	150.533	0.833333	0.24477	0.129642{0	0 0 0 0 0 0 0 0 0 0	6	38	127.833	0.903189	0.951594 -0.07562
	6	6	166	74.2667	0.833333	0.120759	0.139704{0	0 0 0 0 0 0 0 0 0 0	3	28	72.0342	0.897929	0.940417 0.0519889
	6	6	177	73.6	0.833333	0.119675	-0.327946{0	0 0 0 0 0 0 0 0 0 0	4	34	66.1496	0.901874	0.944527 -0.055633
	6	6	204	114.133	0.833333	0.185583	1.80145{0	0 0 1 0 0 1 0 0 0 0	4	32	102.675	0.899902	0.945677 -0.065607
	6	6	497	274.867	0.833333	0.446938	1.71059{0	0 0 0 0 0 0 0 0 0 0	5	38	246.201	0.903847	0.949786 -0.078772
	6	6	391	184.867	0.833333	0.300596	0.51768{0	0 0 0 1 0 0 0 0 0 0	6	37	215.551	0.902532	0.951266 0.0469931
	6	6	233	117.067	0.833333	0.190352	0.956323{0	0 0 0 0 0 0 1 1 0 0	4	31	108.556	0.899902	0.94354 -0.06375
	6	6	360	144.8	0.833333	0.235447	-0.530727{0	0 0 1 0 0 0 0 0 0 0	4	32	105.303	0.900559	0.946006 -0.080075
	5	5	266	111	0.777778	0.180488	-0.306622{1	0 0 0 0 0 0 0 0 0 0	4	36	132.214	0.874754	0.93524 0.059746
	6	6	180	92.0667	0.833333	0.149702	1.08933{0	0 1 0 0 0 0 0 0 0 1	3	34	95.3291	0.902532	0.944855 0.0110298
	6	6	481	229.067	0.833333	0.372466	0.568806{0	0 0 0 0 0 0 0 0 0 0	5	37	212.923	0.903189	0.947321 -0.029641
	6	6	157	57.2667	0.833333	0.093117	-1.08373{0	0 0 0 0 1 0 0 0 0 0	5	31	54.9359	0.896614	0.944034 -0.06093
	6	6	362	174.8	0.833333	0.284228	0.667122{0	0 0 0 0 0 0 0 0 0 0	5	38	166.59	0.903847	0.949786 -0.026226
	6	6	280	117.2	0.833333	0.190569	-0.287703{0	0 0 1 0 0 0 0 0 0 0	6	37	213.538	0.902532	0.951266 0.232335
	6	6	162	63.2	0.833333	0.102764	-0.708286{0	0 0 0 0 0 0 1 0 0 0	6	35	133.821	0.901217	0.950608 0.0994232
	6	6	467	212.333	0.833333	0.345257	0.248456{0	0 0 0 0 0 0 0 0 0 0	6	35	215.03	0.900559	0.950279 -0.031608
	6	6	526	300.133	0.833333	0.488022	1.97153{0	0 0 0 0 0 0 0 0 1 0	6	36	272.103	0.901217	0.950608 -0.021514

Does your data look like this?

Scenario testing (not Bayesian)

Model

.par file, possibly
with free parameters



Simulated
Results

_stat.csv file

Priors?

yes

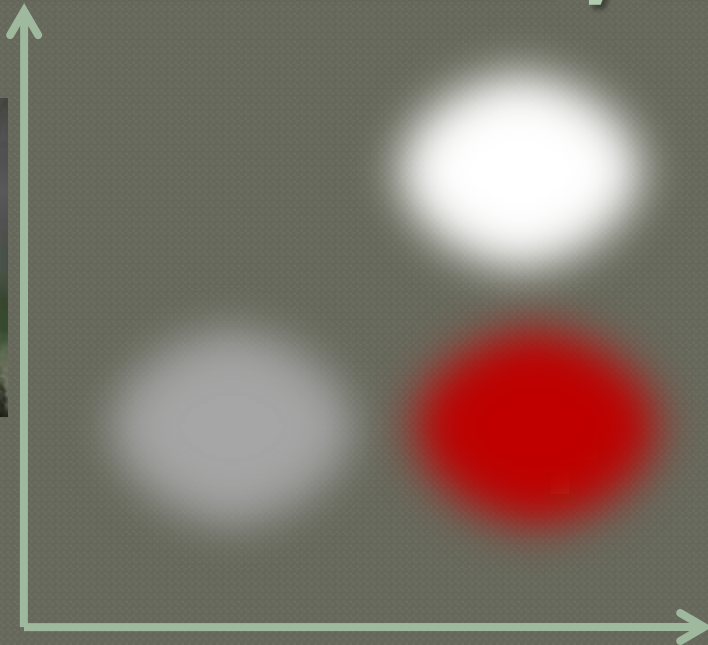
no

AIC/etc

Likelihood!

Scenario testing

Genetic Diversity



H0: The two sites are one large population $N=1000$

H1: The two sites are two small populations $N=500$

H2: A large and a small population $N=750,250$



Scenario testing

PROS

- EXTREMELY EASY to interpret results
- Also easy to implement (both input files and the program)

CONS

- Only allows rejection of hypotheses
- And usually, these are pretty weak hypotheses.
- Can only test a few parameter values
- Not Bayesian (gives probability of data given model parameters, not vice versa).

Bayesian Black Boxes

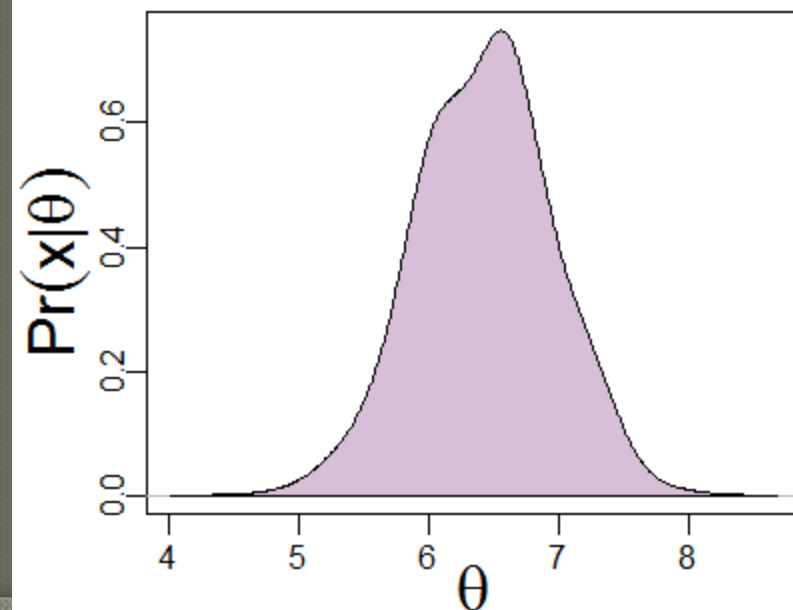
Program



Bayesian

Can θ = some number? Yes or No

Frequentist



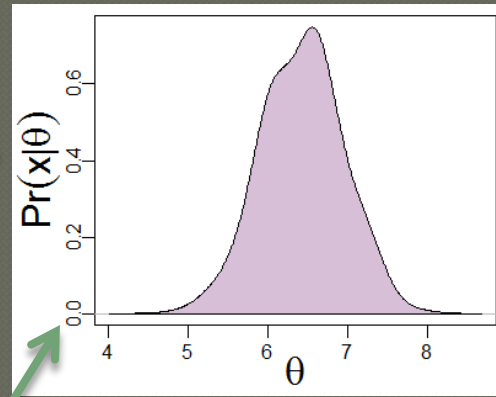
Bayesian simulation-based analysis

Model

.par file, possibly with free parameters



BayeSSC



Posterior Distributions (“Best answers”)

ABC

Rejection or SIR in R; MCMC

Simulated Results

_stat.csv file

Priors?

yes

no

AIC/etc

Likelihood!

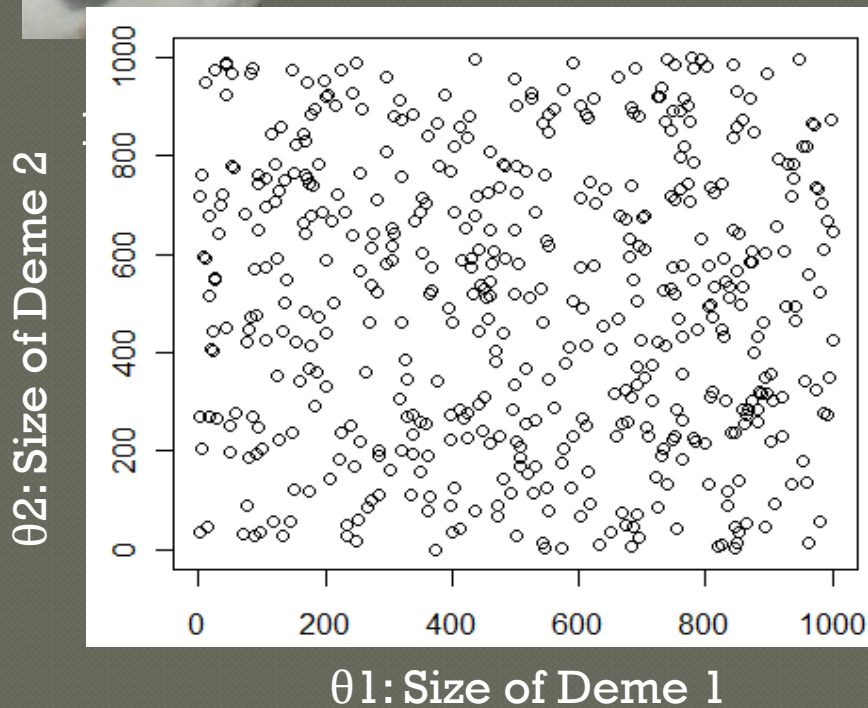
Rejection Method (Pritchard *et al.* 1999)

1. Set up your model
2. Try a lot of different θ values
3. Calculate the difference between your actual data and the simulated data
4. Accept simulations where the difference is sufficiently small

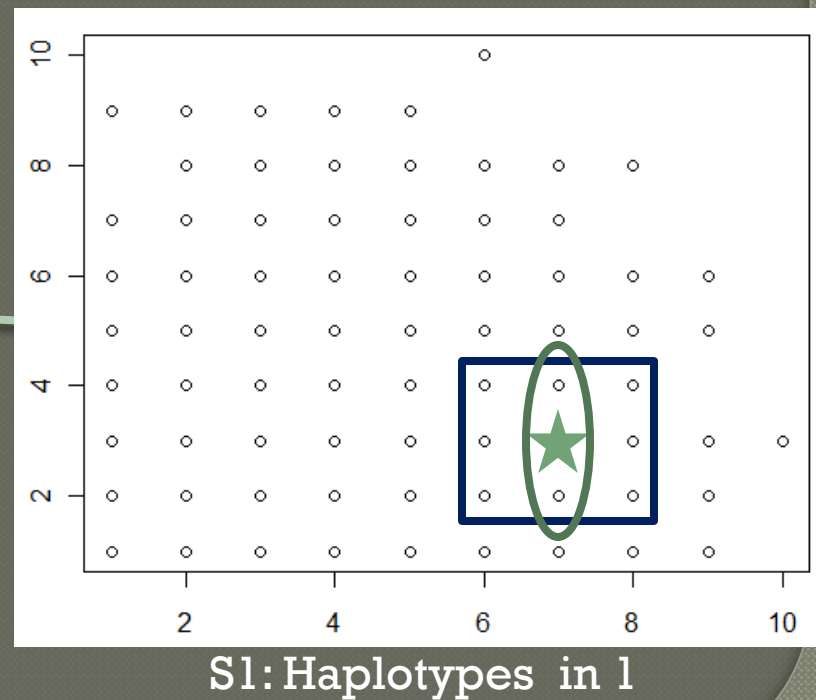
Rejection Method



MODEL: Two populations size θ_1 and θ_2 , both $\sim U(1, 1000)$; split 500ya

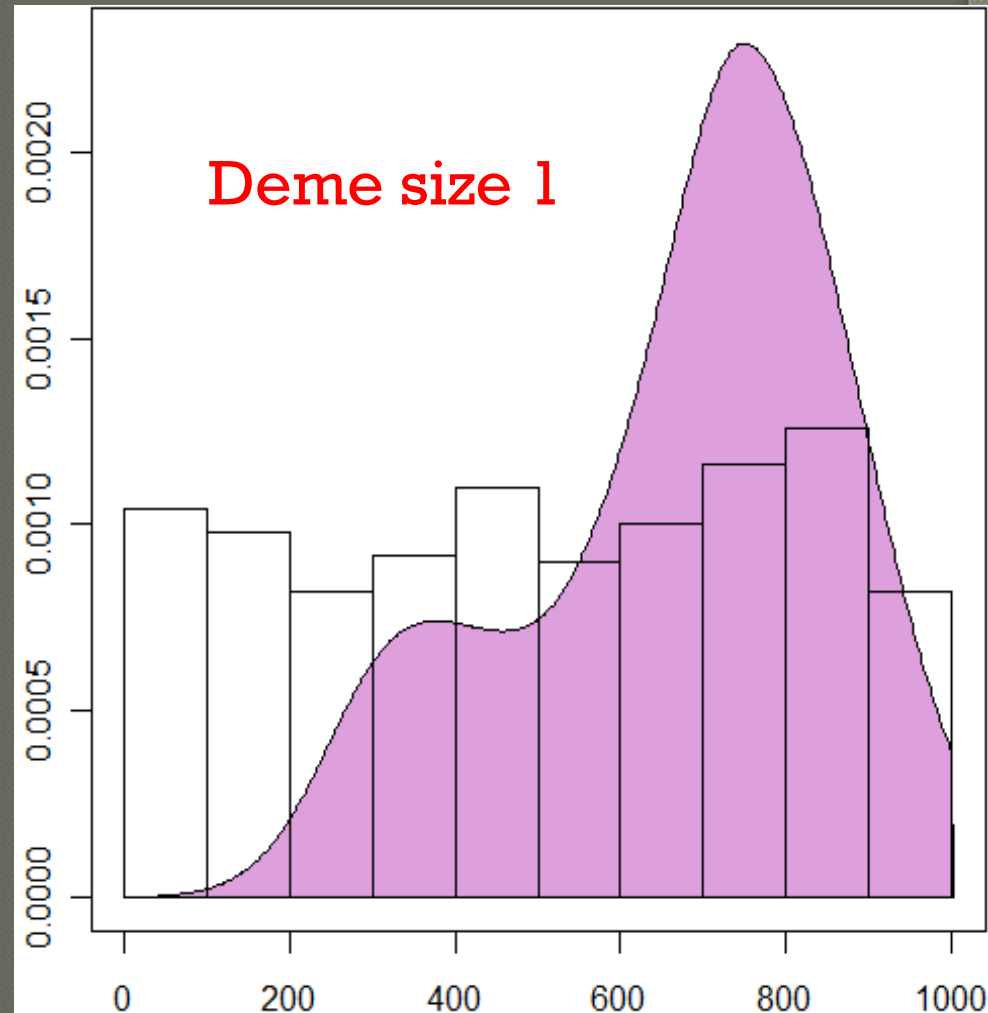


S2: Haplotypes in 2



Rejection method

- Leaves you with a list of “acceptable” parameter values
- Either plot directly
- Or weight by “how” acceptable they were



SIR (Rubin 1988)

◉ Sampling / Importance Resampling

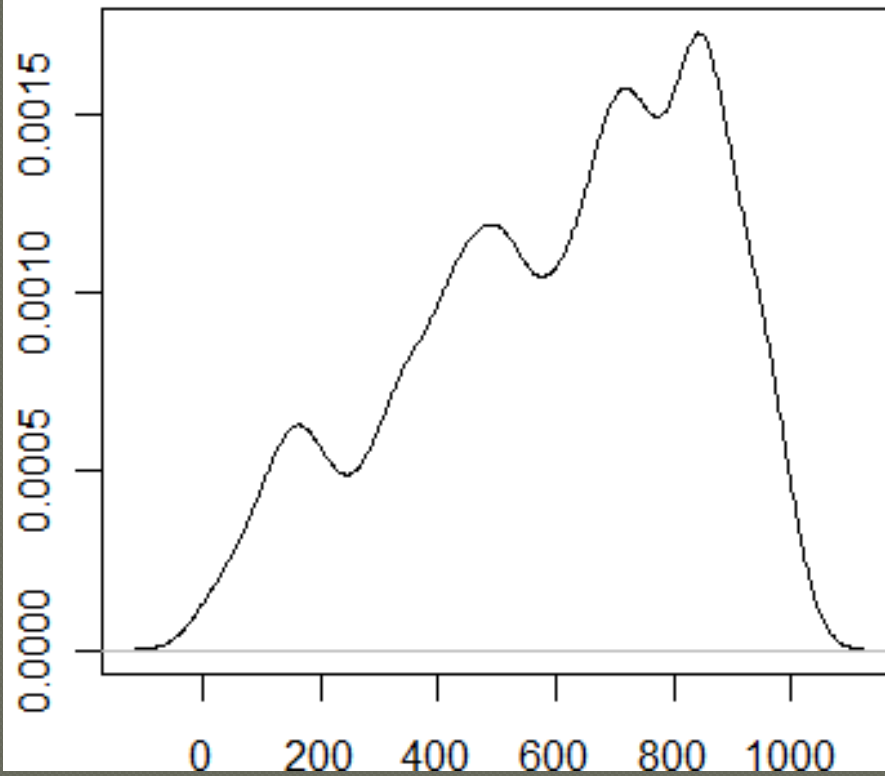
- Generate many values of θ from priors: $g(\theta)$
- Get likelihood of your data for every θ_i : $f(\theta)$
- Weight each point:

$$\omega_i = \frac{f(\theta_i)}{g(\theta_i)}$$

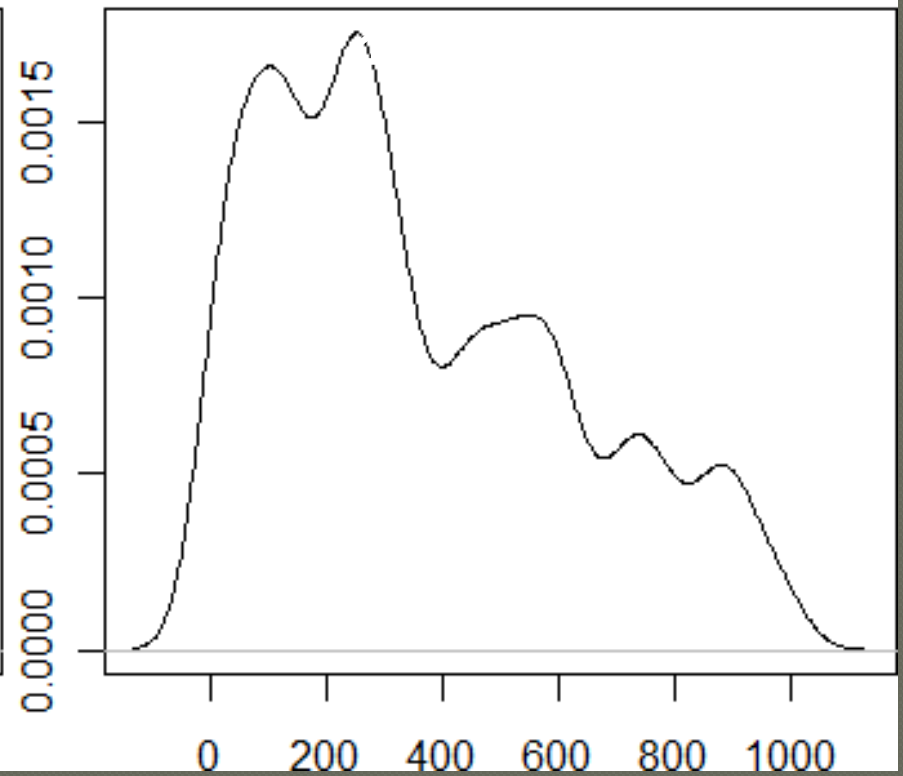
- Resample θ with weight ω

Example

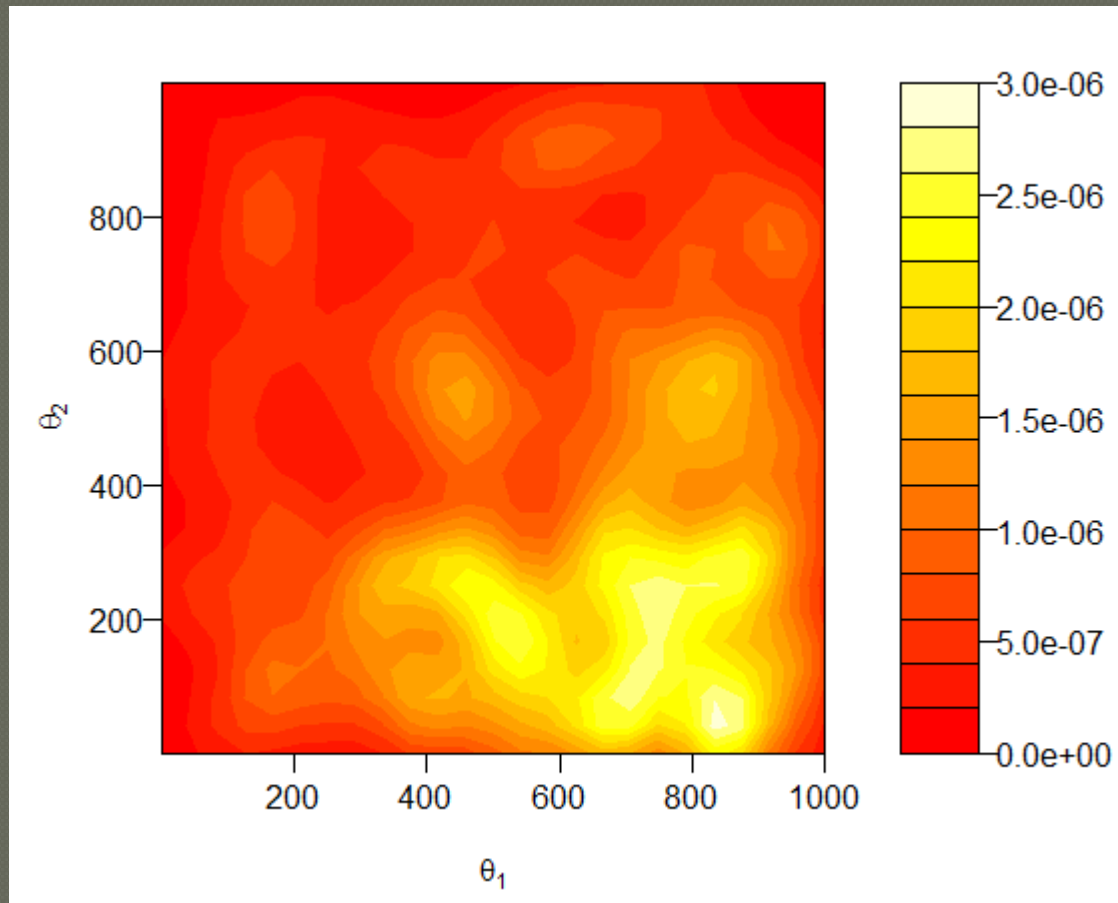
Posterior



Posterior



Example



Model Selection

- Open question as to what constitutes the best method
- Options include AIC, BIC, Bayes Factors, Templeton's NCPA, and Beaumont's head-to-head.

$$AIC = 2p - \text{Log}L(M, D)$$

$$BIC = 2p \log k - \text{Log}L(M, D)$$

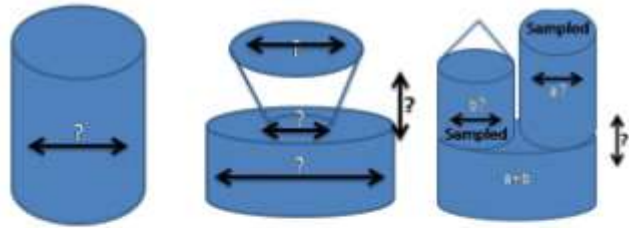
Controversy

- Templeton and Beaumont have been fighting about this in recent years.
- Beaumont, 2008: “As far as we are aware, there are currently no publications other than those of Templeton and co-workers to support the accuracy or efficacy of NCPA.”
- Templeton (PNAS yesterday): “[ABC] is mathematically incorrect and formally illogical.... no statistical or logical validity.”

Likelihood

- Run the program again, with priors drawn from posterior distributions
- Likelihoods are proportional to the number of acceptances of your data
 - Acceptance can be determined in several ways (eg Belle *et al* 2006). We will accept simulations within $x\%$ of the range of values
- Calculate $2p * \log(\% \text{ acceptances})$ for AIC, or $2p * \log(k) * \log(\% \text{ acceptances})$ for BIC

Your turn!



Possible Histories

```
GTTTTACTCACTCT  
H10  
CTAAGCAGGCTCAA  
CCTAACGATTATAG  
TTGAATGTAGATTG  
GTTTTACTCACTCT
```

Sequence Data

