Tutorial BayeSSC

Coalescent methods

A model of evolution that takes into account the uncertainty in genetic polymorphisms simply due to the randomness of mutations and geneologies.

Rosenberg and Nordborg, *Nature Reviews* 2002

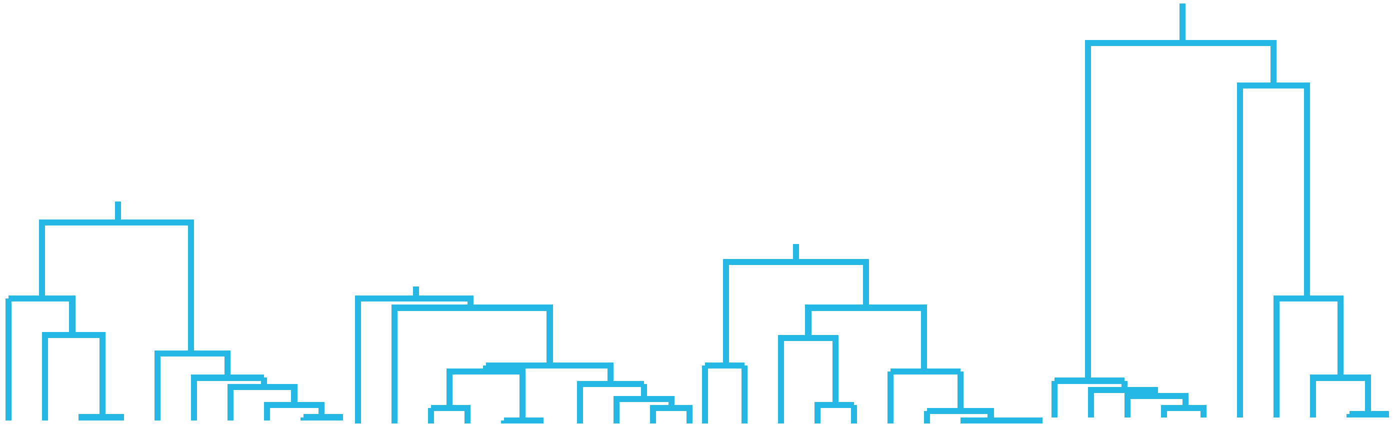
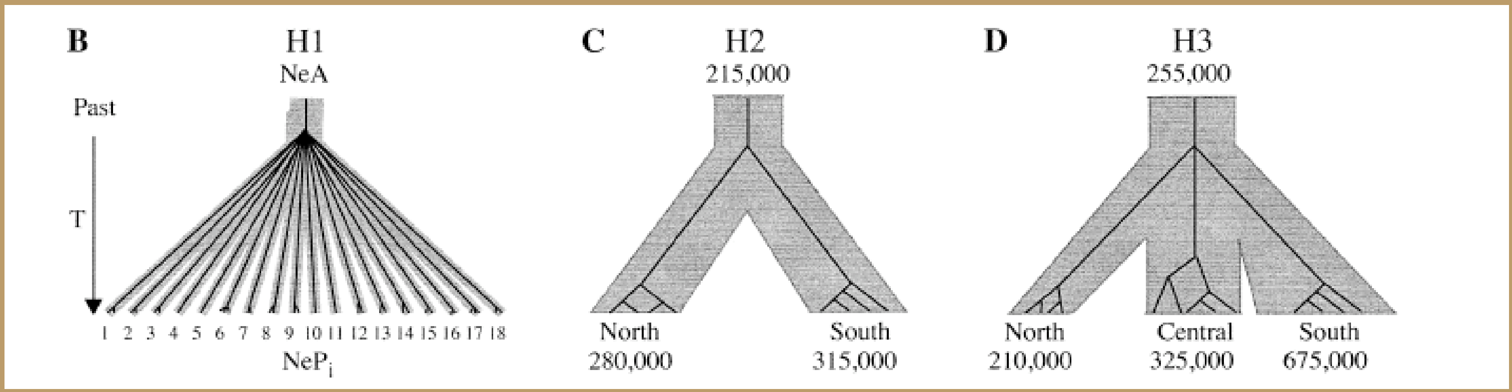


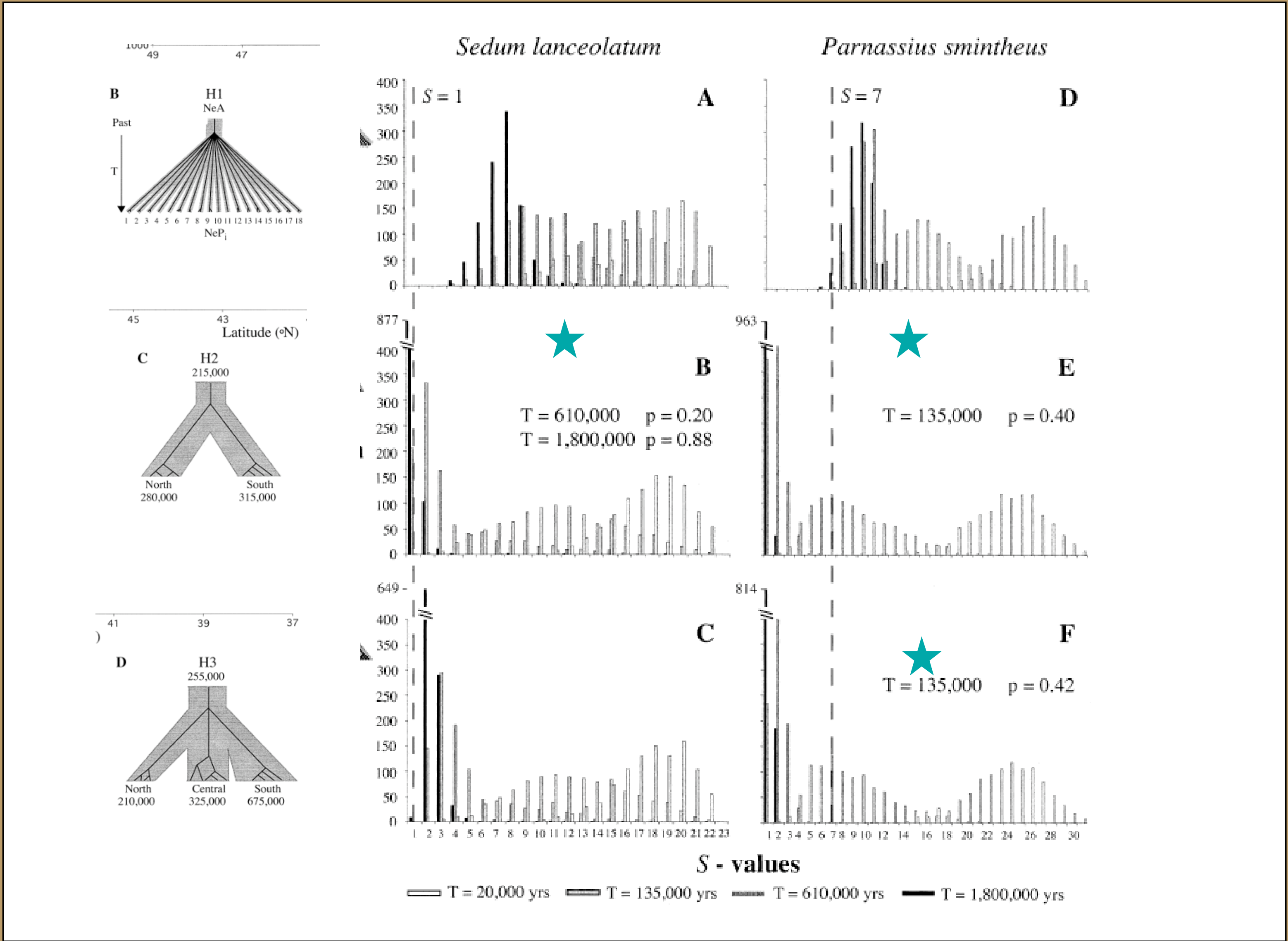
Figure 2 | **Random genealogical trees.** The trees were generated using the same model — the standard coalescent for sample of size ten. Therefore, the variation among the trees reflects chance alone.

Dechaine and Martin 2006

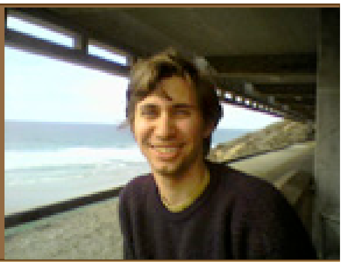
Hypothesis testing framework using data sets simulated under a coalescent model



What hypothesis can you reject?



Serial SimCoal



Coalescent-based hypothesis testing framework with serial sampling

Simulates ancient and modern genetic data for user-specified population histories

What am I doing

Develop a robust method for integrating ancient DNA with other types of data

Right now I am working on integrating Bison ecological niche models with the large ancient DNA data set from Shapiro et al. 2004

Bison Niche Models

Geo-referenced radiocarbon dated bison fossil records

David Nogues-Bravo modeled the climate envelope of bison for several relevant time frames

0kya, 6kya, 18kya, 30kya, 42kya

Bison Potential Distributions

3 Hypotheses produced from 3 different niche models

1 - Hindcasting from current climate data

2 - Hind/forecasting from LGM data

3 - Modeling climate at each temporal layer individually (rarely possible)

Question

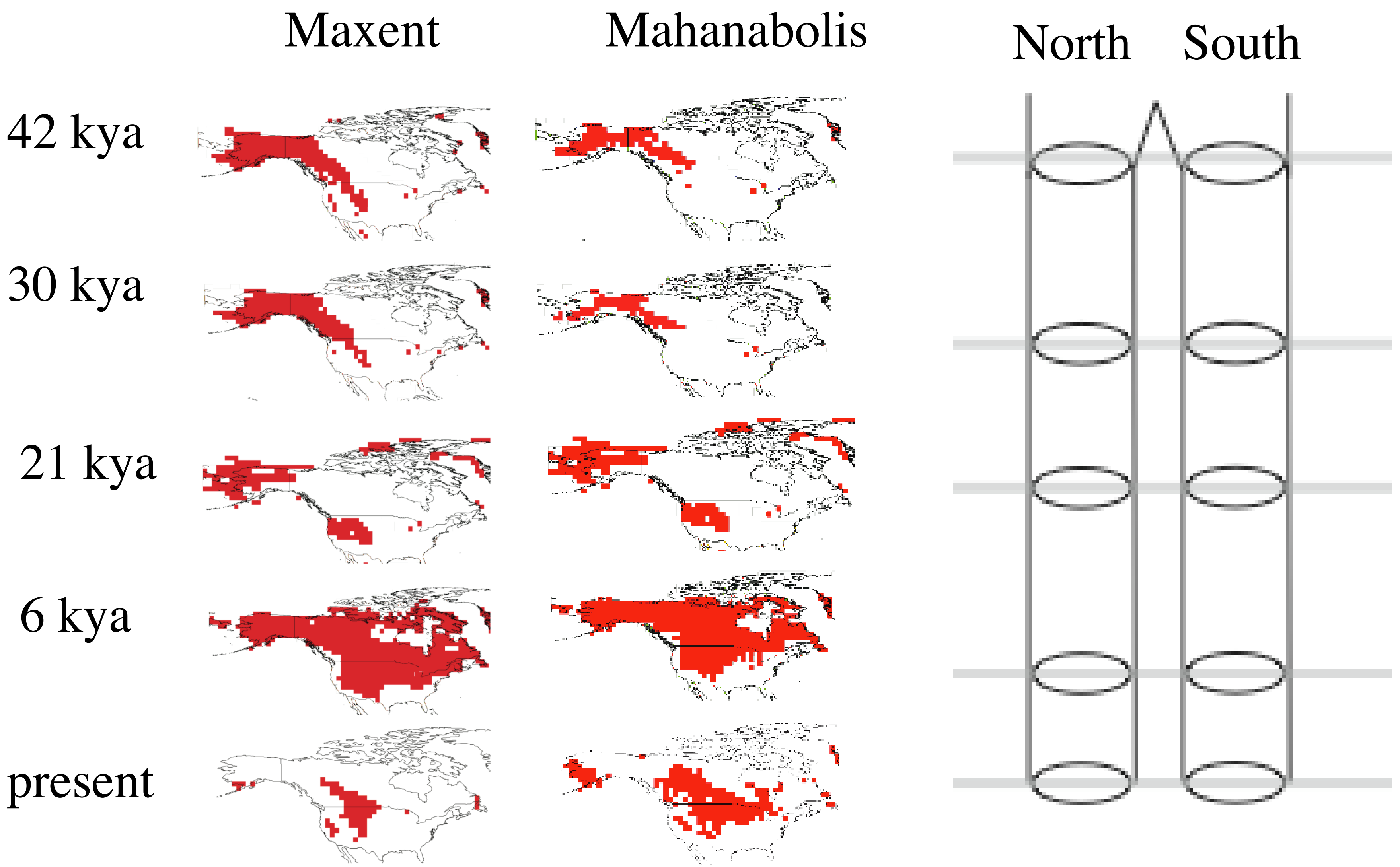
Were bison shifting or tracking their niche over the past 50,000 years?

1 - Hindcasting from current climate data

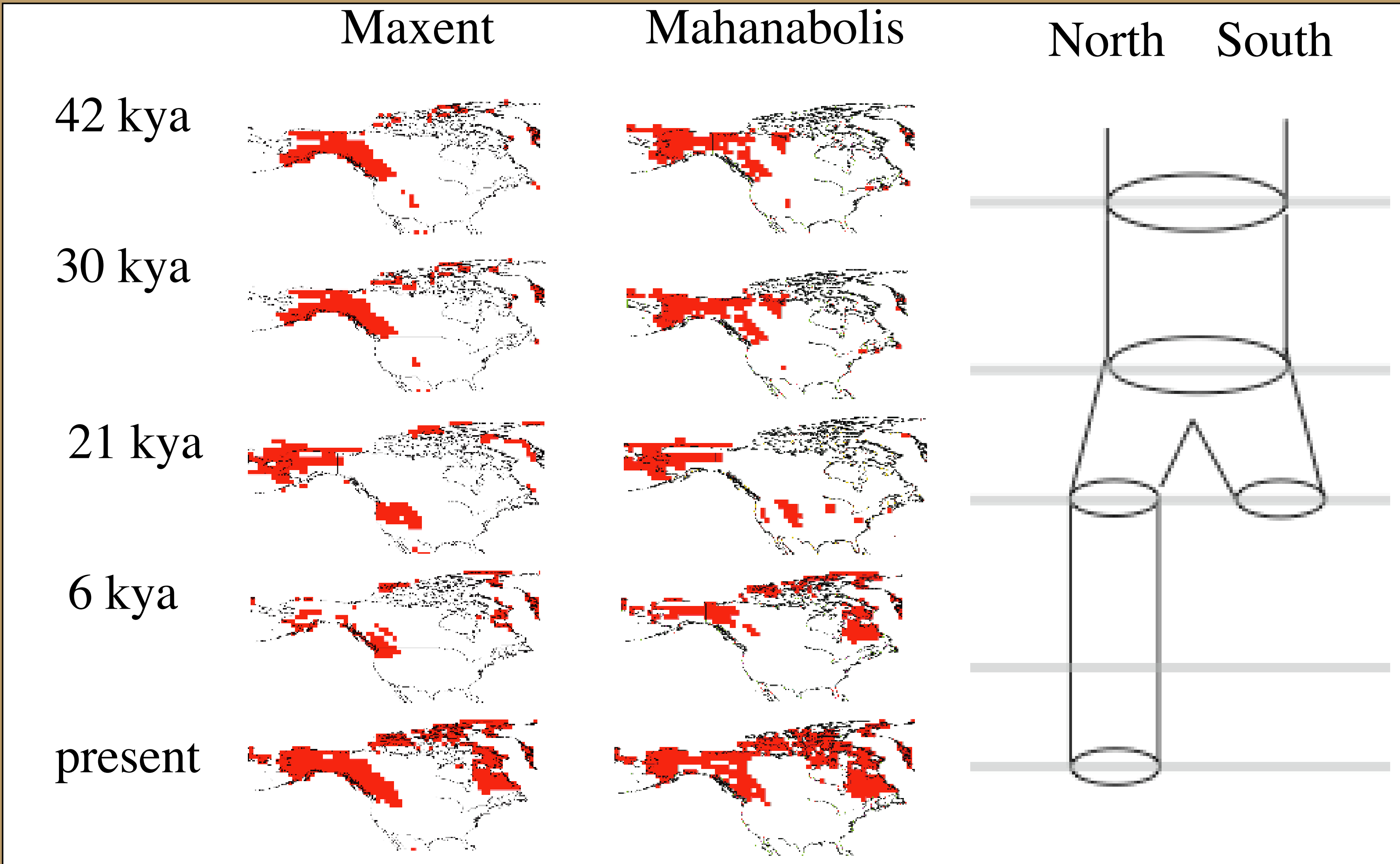
2 - Hind/forecasting from LGM data

3 - Modeling climate at each temporal layer individually (rarely possible)

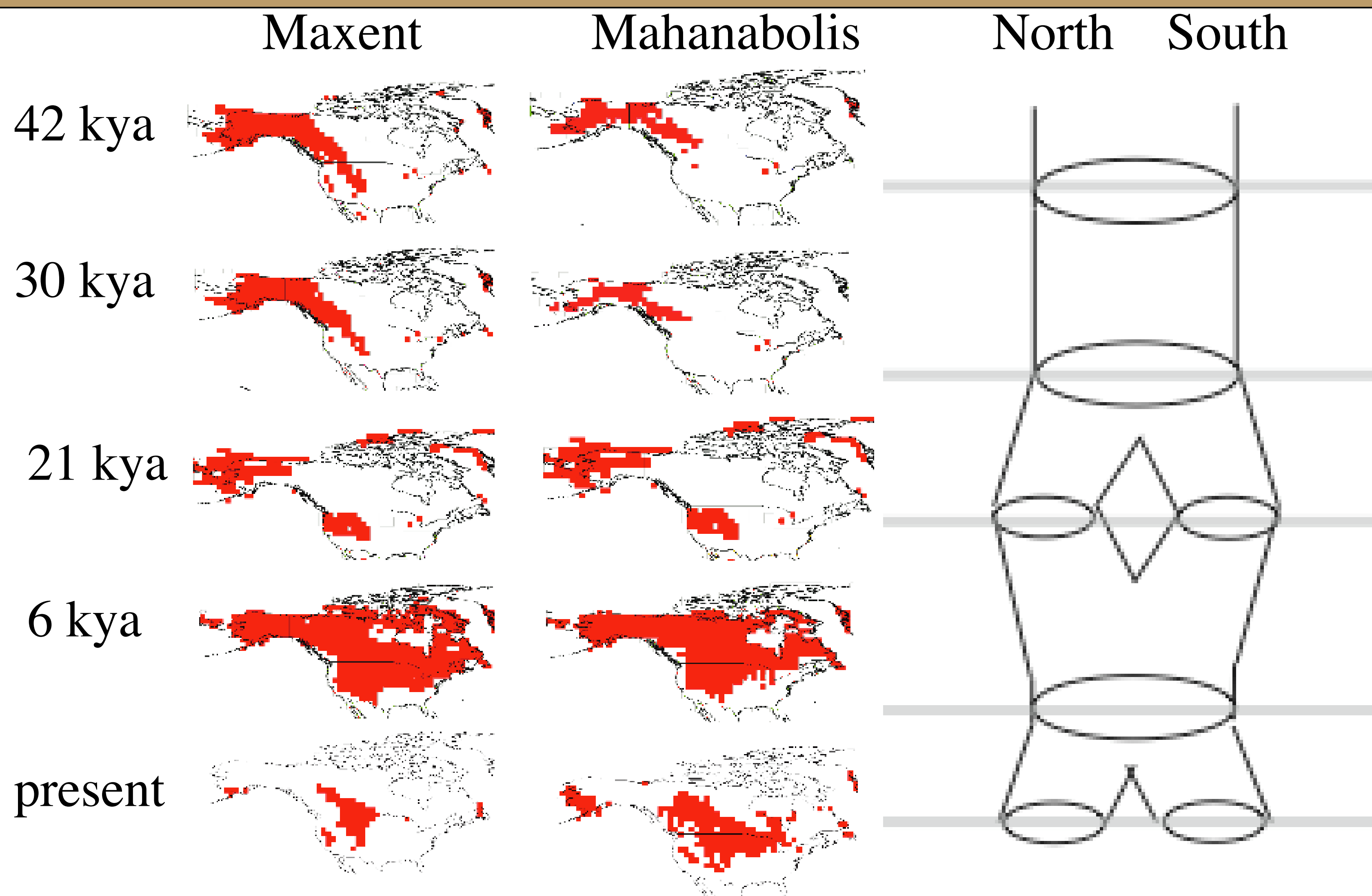
Bison Niche models



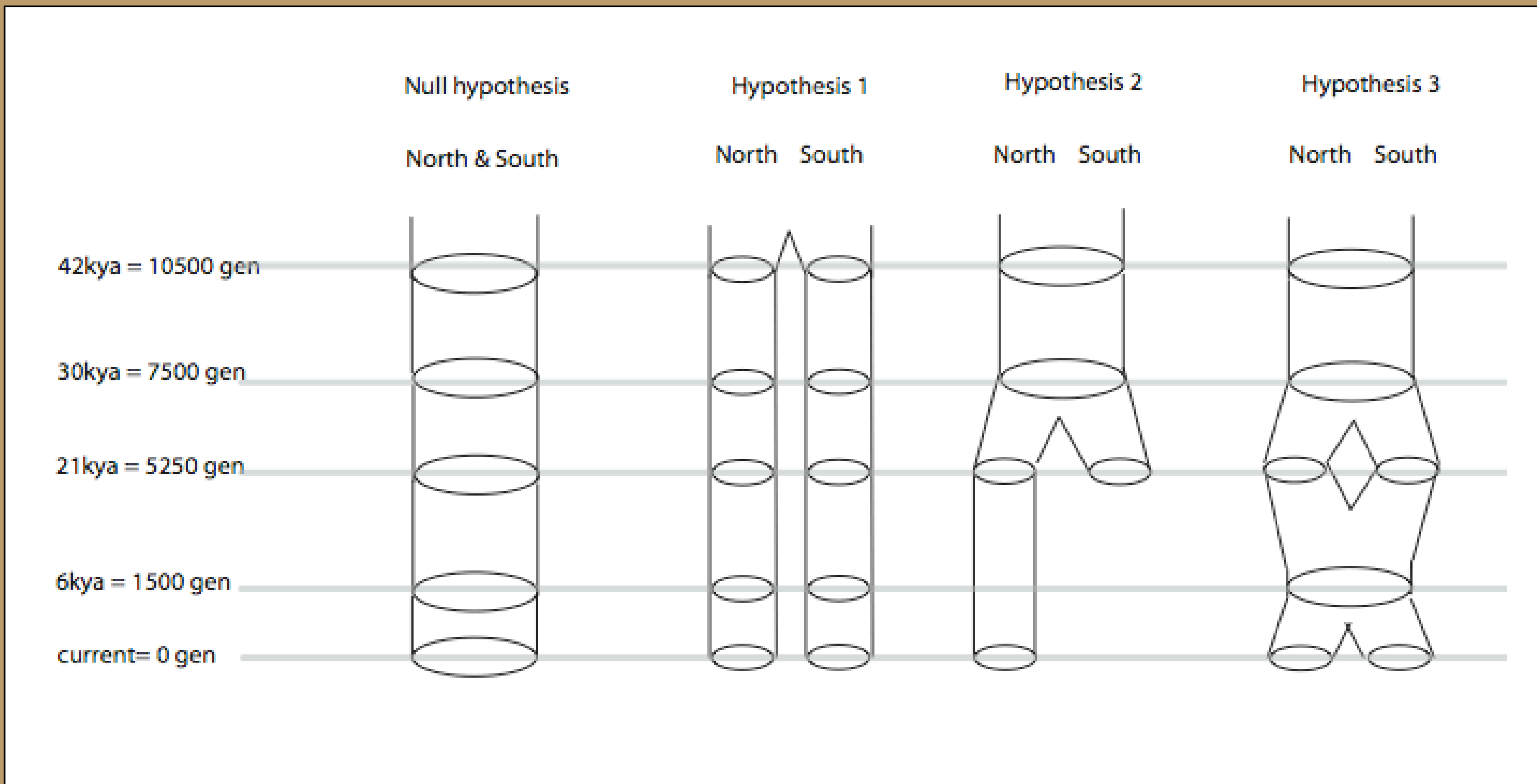
Bison Niche models



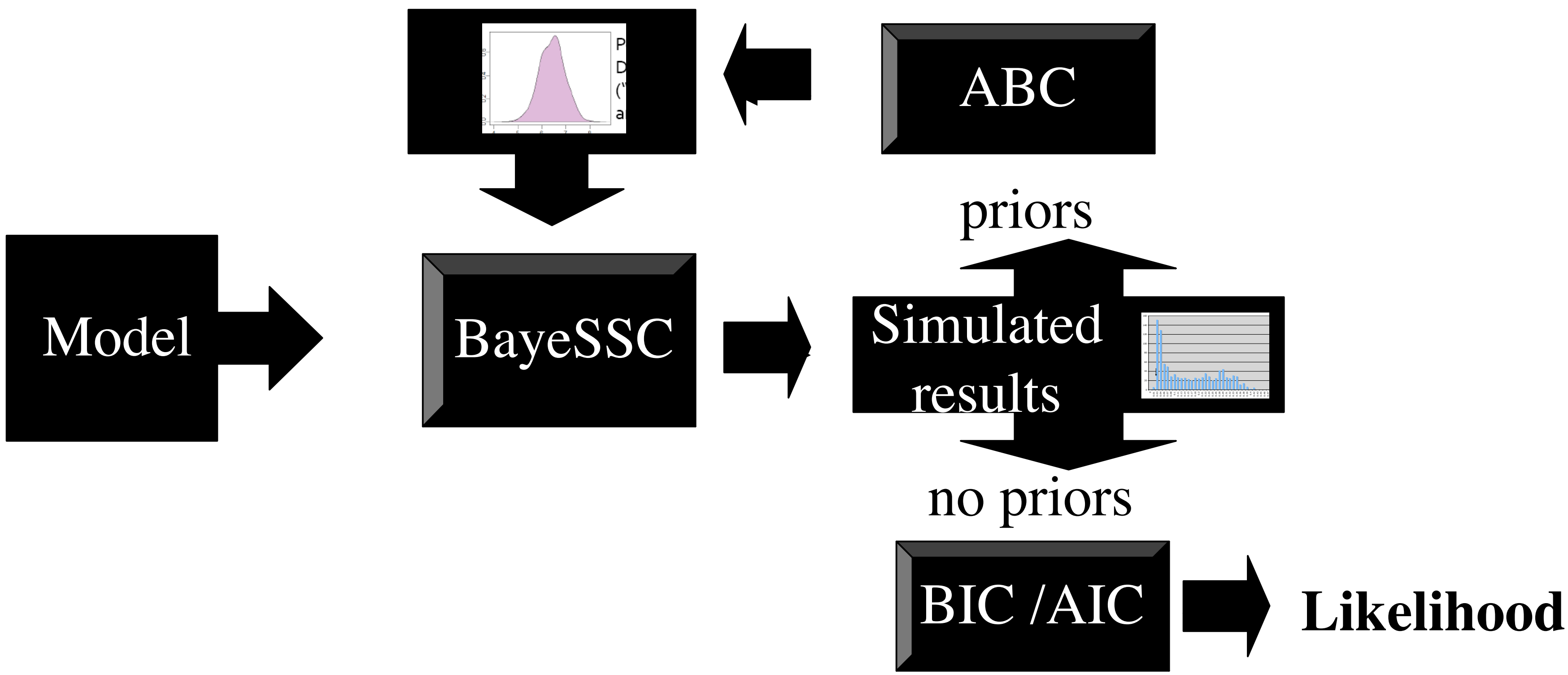
Bison Niche models



Hypothesis Testing



How likely is each model?

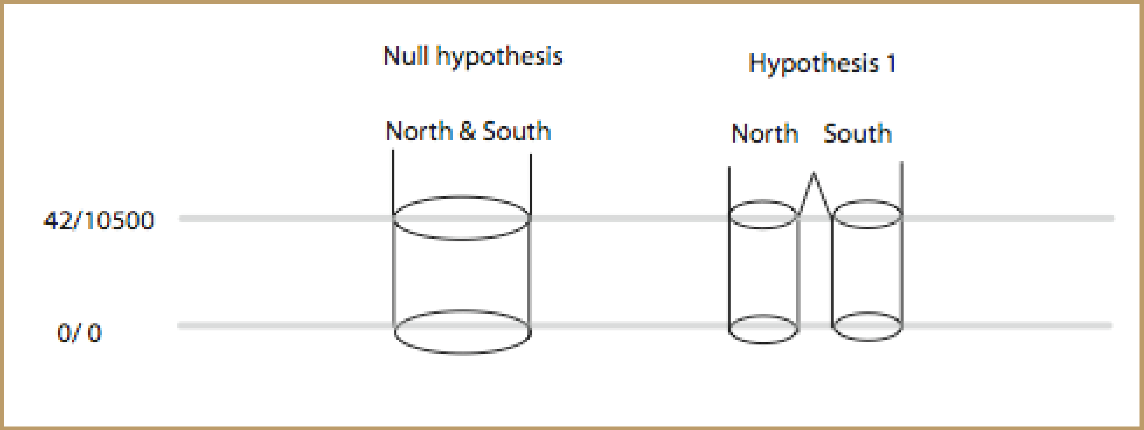


Let's do it!

You need a few things before starting

Your data binned into sample groups in a way that is appropriate to test your hypotheses

A selection of summary statistics from your observed data that are appropriate to test your hypotheses (nucleotide diversity, Fst, etc) using Arlequin or DNAsp



Input File

//Parameters for the coalescence simulation program : simcoal.exe
1 population with ancient DNA
//Population effective sizes (number of genes)
{U:10,1000000}
//Samples sizes:
4 sample groups
6 0 0 0
39 0 0 1
11 12500 0 2
3 12500 0 3
//Growth rates: negative growth implies population expansion
0
Number of migration matrices
0
historical event: time, source, sink, migrants, new deme size, new growth rate, migration matrix index
0 historical event
Mutation rate per generation for the whole sequence
0.0007872
Number of loci
615
//data type either DNA, RFLP, or MICROSAT : If DNA, we need a second term for the transition bias
DNA 0.792
//Gamma parameter (if 0: even mutation rates, if >0 :shape parameter of the Gamma distribution
0

Input File

//Parameters for the coalescence simulation program : simcoal.exe
1 population with ancient DNA
//Population effective sizes (number of genes)
{U:10,1000000}
//Samples sizes:
4 sample groups
6 0 0 0
39 0 0 1
11 12500 0 2
3 12500 0 3
//Growth rates: negative growth implies population expansion
0
Number of migration matrices
0
historical event: time, source, sink, migrants, new deme size, new growth rate, migration matrix index
0 historical event
Mutation rate per generation for the whole sequence
0.0007872
Number of loci
615
//data type either DNA, RFLP, or MICROSAT : If DNA, we need a second term for the transition bias
DNA 0.792
//Gamma parameter (if 0: even mutation rates, if >0 :shape parameter of the Gamma distribution
0

Put the number of demes you wish to simulate after the first comment line. If you are using samples from more than one time point then the text following the number must include the words "with ancient".

Input File

//Parameters for the coalescence simu
1 population with ancient DNA
//Population effective sizes (number
{U:10,1000000}
//Samples sizes:
4 sample groups
6 0 0 0
39 0 0 1
11 12500 0 2
3 12500 0 3
//Growth rates: negative growth impli
0
Number of migration matrices
0
historical event: time, source, sink, m
0 historical event
Mutation rate per generation for the w
0.0007872
Number of loci
615
//data type either DNA, RFLP, or MIC
bias
DNA 0.792
//Gamma parameter (if 0: even mutation rates, if >0 :shape parameter of the Gamma distribution
0

Without ancient information: One sample group per population is assumed. List the number of samples from each population
//Sample Sizes:
20
12
31
With ancient information: An arbitrary number of sampling groups can be added to each population, and they can be pooled together in any combination for statistical analysis. The first line begins with the total number of sampling groups, and can end with any text you want. After that the format is:
First: Number of individuals in sample
Second: Age of the sample (in generations)
Third: The number of the deme the sample belongs to (0,1,2,...)
Fourth: Which stat group the sample group should be pooled with.

Input File

N(t)=N(0)ert

Enter one value per population. Because coalescent simulations run backward through time, a negative growth rate implies a population larger now than in the past.

Example: Two stable populations, and one that is growing 2% per generation

//Growth rates:

0

0

-.02

//Parameters for the coa

1 population with anc

//Population effective si

{U:10,1000000}

//Samples sizes:

4 sample groups

6 0 0 0

39 0 0 1

11 12500 0 2

3 12500 0 3

//Growth rates: negative growth implies population expansion

0

Number of migration matrices

0

historical event: time, source, sink, migrants, new deme size, new growth rate, migration matrix index

0 historical event

Mutation rate per generation for the whole sequence

0.0007872

Number of loci

615

//data type either DNA, RFLP, or MICROSAT : If DNA, we need a second term for the transition bias

DNA 0.792

//Gamma parameter (if 0: even mutation rates, if >0 :shape parameter of the Gamma distribution

0

Input File

//Parameters for the coalescence si
1 population with ancient DNA
//Population effective sizes (numbe
{U:10,1000000}
//Samples sizes:
4 sample groups
6 0 0 0
39 0 0 1
11 12500 0 2
3 12500 0 3
//Growth rates: negative growth im
0
Number of migration matrices
0
historical event: time, source, sink,
0 historical event
Mutation rate per generation for the
0.0007872
Number of loci
615
//data type either DNA, RFLP, or MICROSAT : If DNA, we need a second term for the transition
bias
DNA 0.792
//Gamma parameter (if 0: even mutation rates, if >0 :shape parameter of the Gamma distribution
0

The first line begins with the number of matrices (0 is fine). The next lines define the ratio of migrants from each deme to each deme; each migration matrix must be preceeded by a comment. The first migration matrix is assumed to represent the migration in the present (or at t=0). If you have more than one population but no migration, then the demes will NEVER coalesce and you will get no information. Note that the diagonal elements of the matrix are meaningless, but the simulations will run faster if you set them to 0.

//Migration matrices
2
//Matrix 0: Deme0 <-> Deme1 <-> Deme2
0 .01 0
.01 0 .01
0 .01 0
//Matrix 1: Migration stopped
0 0 0
0 0 0
0 0 0

Input File

1. The time (in generations) when the event occured
2. The source deme (0,1,2...)
3. The sink deme.
4. The proportion of the source that migrates to the sink. It also represents the probability for each lineage in the source deme to migrate in the sink deme. If no migration is involved in the event, then just specify the same source, sink, and a migration probability of 0.
5. The new effective population size of the sink deme relative to one generation later in time. Remeber, coalescent simulations run backwards. So a value of 0.5 here implies the event doubled the population (think, "The population used to be half as big").
6. The new growth rate of the sink deme. Negative values mean the population is growing.
7. The id of the new migration matrix to use for all demes.

Example: 2000 generations ago, deme 0 and 2 split from what used to be a larger deme 1
//Format: time, src, sink, % mig, new Nef, new r, MigMat
2000 0 1 1 2 0 1
2000 2 1 1 1 0 1

historical event: time, source, sink, migrants, new deme size, new growth rate, migration matrix index
0 historical event
Mutation rate per generation for the whole sequence
0.0007872
Number of loci
615
//data type either DNA, RFLP, or MICROSAT : If DNA, we need a second term for the transition bias
DNA 0.792
//Gamma parameter (if 0: even mutation rates, if >0 :shape parameter of the Gamma distribution
0

Input File

//Parameters for the coalescence simulation program : simcoal.exe
1 population with ancient DNA
//Populat
{U:10,10
//Sample
4 sample
6 0 0 0
39 0 0 1
11 12500
3 12500 0 5

average mutation number of mutations per generation per nucleotide, times the number of nucleotides.
Example: 10%/bp/million years for a 300bp sequence and a species whose generations are 5 years long
10%/bp/1,000,000yr = .00000001/bp/yr * 300bp = .00003/yr * 5 yr/gen = .00015/gen
//Mutation rate
.00015

//Growth rates: negative growth implies population expansion
0
Number of migration matrices
0
historical event: time, source, sink, migrants, new deme size, new growth rate, migration matrix index
0 historical event
Mutation rate per generation for the whole sequence
0.0007872
Number of loci
615
//data type either DNA, RFLP, or MICROSAT : If DNA, we need a second term for the transition bias
DNA 0.792
//Gamma parameter (if 0: even mutation rates, if >0 :shape parameter of the Gamma distribution
0

Input File

//Parameters for the coalescence simulation program : simcoal.exe
1 population with ancient DNA
//Population effective sizes (number of genes)
{U:10,1000000}
//Samples sizes:
4 sample groups
6 0 0 0
39 0 0 1
11 12500 0 2
3 12500 0 3
//Growth rates: negative growth implies population expansion
0
Number of migration matrices
0

For DNA, the length of the sequence to simulate. For RFLP and STRs, the number of RFLPs/STRs to simulate.

historical e
0 historical
Mutation rate per generation for the whole sequence
0.0007872
Number of loci
615
//data type either DNA, RFLP, or MICROSAT : If DNA, we need a second term for the transition bias
DNA 0.792
//Gamma parameter (if 0: even mutation rates, if >0 :shape parameter of the Gamma distribution
0

Input File

//Parameters for the coalescence simulation program : simcoal.exe
1 popul
//Popula
{U:10,1
//Sampl
4 sampl
6 0 0 0
39 0 0 1
11 1250
3 12500
//Growt
0
Numbe
0

Type of data to simulate:

* MICROSAT: Microsatellites are simulated with a pure stepwise model, and can be followed with a range constraint if you wish (no number implies no limit).
* DNA: followed by the transition/transversion bias number. Mutation probabilities can be heterogenous (see "Gamma")
* RFLP: a two allele model.

Example 1: Using DNA where 1/3 of the mutations are A<->G or C<->T (all mutations are equally likely)
//Number of loci:
DNA 0.33333

historical event: time, source, sink, migrants, new deme size, new growth rate, migration matrix index
0 historical event
Mutation rate per generation for the whole sequence
0.0007872
Number of loci
615
//data type either DNA, RFLP, or MICROSAT : If DNA, we need a second term for the transition bias
DNA 0.792
//Gamma parameter (if 0: even mutation rates, if >0 :shape parameter of the Gamma distribution
0

Input File

//Parameters for the coalescence simulation program : simcoal.exe
1 population with ancient DNA
//Population effective sizes (number of genes)
{U:10,1000000}
//Samples sizes:
4 sample
6 0 0 0
39 0 0 1
11 12500
3 12500
//Growth
0
Number
0
historica ... k
0 historic
Mutation
0.000787
Number
615
//data type either DNA, RFLP, or MICROSAT : If DNA, we need a second term for the transition bias
DNA 0.792
//Gamma parameter (if 0: even mutation rates, if >0 :shape parameter of the Gamma distribution
0

hese parameters control the heterogeneity of DNA mutation rates along the sequence. The first number is the shape parameter a of a Gamma distribution of mutation rates. If a value of zero is entered, then an even mutation rate model is implemented. The second number is the number of rate classes to simulate. If a value of zero is entered, then a continuous distribution is used (as many classes as there are loci or nucleotides).

Example 1: Uniform mutation rates (Cantor-Jukes model)
//Gamma distribution for mutation:
0 0

Example 2: Heterogenous mutation (Kimura 2-Parameter model)
//Gamma distribution for mutation:
0.4 10

Let’s run the file

Double click the BayeSSC program

Make sure the input file has a .par extenstion

Type the name of the input file without the .par

Choose number of samples to simulate (let’s do 1000)

Open .csv file

Columns of summary statistics

GROUP 0	Haptypes	PrivHaps	SegSites	PairDiffs	HapDiver	NucltdDiv	TajimasD	MismatDist	0 VS 1	PrivTo0	PrivTo1	PairDiffs	MeanDiv(H	PoolDiv(Ht)	Fst
	1	1	0	0	0	0	nan	{15 }		0	0	0	0	0	nan
	6	6	431	192.267	0.833333	0.312629	0.12092	{0 0 0 0 0 0 0 0 1 0 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 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 0 1 0 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 0 1 0		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 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 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 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 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 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 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 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 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 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 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 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 0 1		4	32	105.303	0.900559	0.946006	-0.080079
	5	5	266	111	0.777778	0.180488	-0.306622	{1 0 0 0 0 0 0 0 0 0 0 0 2		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 0 0		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 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 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 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 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 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 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 0 0		6	36	272.103	0.901217	0.950608	-0.021514

Open .csv file

Columns of summary statistics

GROUP 0	Haptypes	PrivHaps	SegSites	PairDiffs	HapDiver	NucltdDiv	TajimasD	MismatDist	0 VS 1	PrivTo0	PrivTo1	PairDiffs	MeanDiv(H	PoolDiv(Ht)	Fst
	1	1	0	0	0	0	nan	{15 }		0	0	0	0	0	nan
	6	6	431	192.267	0.833333	0.312629	0.12092	{0 0 0 0 0 0 0 0 1 0 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 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 0 1 0 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 0 1 0		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 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 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 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 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 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 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 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 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 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 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 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 0 1		4	32	105.303	0.900559	0.946006	-0.080079
	5	5	266	111	0.777778	0.180488	-0.306622	{1 0 0 0 0 0 0 0 0 0 0 0 2		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 0 0		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 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 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 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 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 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 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 0 0		6	36	272.103	0.901217	0.950608	-0.021514

Does your data look like this?

Using R to estimate posteriors

Double click R

Copy and paste text from eval.r

Type
reject(”directorypathfilename_stat.csv”)

You should get a list of column numbers and stats

Using R to estimate posteriors

```
> reject("/Users/jessica/Tutorial/tutorialnullk500k/tutorialnull_stat.csv")
Loading required package: locfit
Loading required package: akima
Loading required package: lattice
locfit 1.5-4     2007-11-27
 [1] "GROUP.0"        "Haptypes"       "PrivHaps"
 [4] "SegSites"       "PairDiffs"      "HapDiver"
 [7] "NucltdDiv"      "TajimasD"       "MismatDist"
[10] "X0.VS.1"        "PrivTo0"        "PrivTo1"
[13] "PairDiffs.1"    "MeanDiv.Hs.bar." "PoolDiv.Ht."
[16] "Fst"            "X0.VS.2"        "PrivTo0.1"
[19] "PrivTo2"        "PairDiffs.2"    "MeanDiv.Hs.bar..1"
[22] "PoolDiv.Ht..1"  "Fst.1"          "X0.VS.3"
[25] "PrivTo0.2"      "PrivTo3"        "PairDiffs.3"
[28] "MeanDiv.Hs.bar..2" "PoolDiv.Ht..2"  "Fst.2"
[31] "GROUP.1"        "Haptypes.1"     "PrivHaps.1"
[34] "SegSites.1"     "PairDiffs.4"    "HapDiver.1"
[37] "NucltdDiv.1"    "TajimasD.1"     "MismatDist.1"
[40] "X1.VS.2"        "PrivTo1.1"      "PrivTo2.1"
[43] "PairDiffs.5"    "MeanDiv.Hs.bar..3" "PoolDiv.Ht..3"
[46] "Fst.3"          "X1.VS.3"        "PrivTo1.2"
[49] "PrivTo3.1"      "PairDiffs.6"    "MeanDiv.Hs.bar..4"
[52] "PoolDiv.Ht..4"  "Fst.4"          "GROUP.2"
[55] "Haptypes.2"     "PrivHaps.2"     "SegSites.2"
[58] "PairDiffs.7"    "HapDiver.2"     "NucltdDiv.2"
[61] "TajimasD.2"     "MismatDist.2"   "X2.VS.3"
[64] "PrivTo2.2"      "PrivTo3.2"      "PairDiffs.8"
[67] "MeanDiv.Hs.bar..5" "PoolDiv.Ht..5"  "Fst.5"
[70] "GROUP.3"        "Haptypes.3"     "PrivHaps.3"
[73] "SegSites.3"     "PairDiffs.9"    "HapDiver.3"
[76] "NucltdDiv.3"    "TajimasD.3"     "MismatDist.3"
[79] "COMBINED"       "Haptypes.4"     "PrivHaps.4"
[82] "SegSites.4"     "PairDiffs.10"   "HapDiver.4"
[85] "NucltdDiv.4"    "TajimasD.4"     "MismatDist.4"
[88] "PRIORS"         "Deme.Size.0"
Which column/s (eg 4,23,27)?
```

Using R to estimate posteriors

```
> reject("/Users/jessica/Tutorial/tutorialnullk500k/tutorialnull_stat.csv")
Loading required package: locfit
Loading required package: akima
Loading required package: lattice
locfit 1.5-4     2007-11-27
 [1] "GROUP.0"        "Haptypes"       "PrivHaps"
 [4] "SegSites"       "PairDiffs"      "HapDiver"
 [7] "NucltdDiv"      "TajimasD"       "MismatDist"
[10] "X0.VS.1"        "PrivTo0"        "PrivTo1"
[13] "PairDiffs.1"    "MeanDiv.Hs.bar." "PoolDiv.Ht."
[16] "Fst"            "X0.VS.2"        "PrivTo0.1"
[19] "PrivTo2"        "PairDiffs.2"    "MeanDiv.Hs.bar..1"
[22] "PoolDiv.Ht..1"  "Fst.1"          "X0.VS.3"
[25] "PrivTo0.2"      "PrivTo3"        "PairDiffs.3"
[28] "MeanDiv.Hs.bar..2" "PoolDiv.Ht..2"  "Fst.2"
[31] "GROUP.1"        "Haptypes.1"     "PrivHaps.1"
[34] "SegSites.1"     "PairDiffs.4"    "HapDiver.1"
[37] "NucltdDiv.1"    "TajimasD.1"     "MismatDist.1"
[40] "X1.VS.2"        "PrivTo1.1"      "PrivTo2.1"
[43] "PairDiffs.5"    "MeanDiv.Hs.bar..3" "PoolDiv.Ht..3"
[46] "Fst.3"          "X1.VS.3"        "PrivTo1.2"
[49] "PrivTo3.1"      "PairDiffs.6"    "MeanDiv.Hs.bar..4"
[52] "PoolDiv.Ht..4"  "Fst.4"          "GROUP.2"
[55] "Haptypes.2"     "PrivHaps.2"     "SegSites.2"
[58] "PairDiffs.7"    "HapDiver.2"     "NucltdDiv.2"
[61] "TajimasD.2"     "MismatDist.2"   "X2.VS.3"
[64] "PrivTo2.2"      "PrivTo3.2"      "PairDiffs.8"
[67] "MeanDiv.Hs.bar..5" "PoolDiv.Ht..5"  "Fst.5"
[70] "GROUP.3"        "Haptypes.3"     "PrivHaps.3"
[73] "SegSites.3"     "PairDiffs.9"    "HapDiver.3"
[76] "NucltdDiv.3"    "TajimasD.3"     "MismatDist.3"
[79] "COMBINED"       "Haptypes.4"     "PrivHaps.4"
[82] "SegSites.4"     "PairDiffs.10"   "HapDiver.4"
[85] "NucltdDiv.4"    "TajimasD.4"     "MismatDist.4"
[88] "PRIORS"         "Deme.Size.0"
Which column/s (eg 4,23,27)?
```

Statistic	.csv column	Observed value
NuclDiv0	7	0.01931
NuclDiv1	37	0.0116
NuclDiv2	60	0.024
NuclDiv3	76	0.035
Fst 0v1	16	0.61
Fst 0v2	69	0.00063

Using R to estimate posteriors

```
> reject("/Users/jessica/Tutorial/tutorialnullk500k/tutorialnull_stat.csv")
Loading required package: locfit
Loading required package: akima
Loading required package: lattice
locfit 1.5-4     2007-11-27
 [1] "GROUP.0"        "Haptypes"       "PrivHaps"
 [4] "SegSites"       "PairDiffs"      "HapDiver"
 [7] "NucltdDiv"      "TajimasD"       "MismatDist"
[10] "X0.VS.1"        "PrivTo0"        "PrivTo1"
[13] "PairDiffs.1"    "MeanDiv.Hs.bar." "PoolDiv.Ht."
[16] "Fst"            "X0.VS.2"        "PrivTo0.1"
[19] "PrivTo2"        "PairDiffs.2"    "MeanDiv.Hs.bar..1"
[22] "PoolDiv.Ht..1"  "Fst.1"          "X0.VS.3"
[25] "PrivTo0.2"      "PrivTo3"        "PairDiffs.3"
[28] "MeanDiv.Hs.bar..2" "PoolDiv.Ht..2" "Fst.2"
[31] "GROUP.1"        "Haptypes.1"     "PrivHaps.1"
[34] "SegSites.1"     "PairDiffs.4"    "HapDiver.1"
[37] "NucltdDiv.1"    "TajimasD.1"     "MismatDist.1"
[40] "X1.VS.2"        "PrivTo1.1"      "PrivTo2.1"
[43] "PairDiffs.5"    "MeanDiv.Hs.bar..3" "PoolDiv.Ht..3"
[46] "Fst.3"          "X1.VS.3"        "PrivTo1.2"
[49] "PrivTo3.1"      "PairDiffs.6"    "MeanDiv.Hs.bar..4"
[52] "PoolDiv.Ht..4"  "Fst.4"          "GROUP.2"
[55] "Haptypes.2"     "PrivHaps.2"     "SegSites.2"
[58] "PairDiffs.7"    "HapDiver.2"     "NucltdDiv.2"
[61] "TajimasD.2"     "MismatDist.2"   "X2.VS.3"
[64] "PrivTo2.2"      "PrivTo3.2"      "PairDiffs.8"
[67] "MeanDiv.Hs.bar..5" "PoolDiv.Ht..5" "Fst.5"
[70] "GROUP.3"        "Haptypes.3"     "PrivHaps.3"
[73] "SegSites.3"     "PairDiffs.9"    "HapDiver.3"
[76] "NucltdDiv.3"    "TajimasD.3"     "MismatDist.3"
[79] "COMBINED"       "Haptypes.4"     "PrivHaps.4"
[82] "SegSites.4"     "PairDiffs.10"   "HapDiver.4"
[85] "NucltdDiv.4"    "TajimasD.4"     "MismatDist.4"
[88] "PRIORS"         "Deme.Size.0"
Which column/s (eg 4,23,27)? 7,37,60,76,16,69
Observed values:  0.01931,0.0116,0.024,0.035,0.61,0.00063
```

Statistic	.csv column	Observed value
NuclDiv0	7	0.01931
NuclDiv1	37	0.0116
NuclDiv2	60	0.024
NuclDiv3	76	0.035
Fst 0v1	16	0.61
Fst 0v2	69	0.00063

Using R to estimate posteriors

0.1%	1%	5%	10%	25%
0.9566305	1.2564563	1.4953178	1.6186482	1.7954118

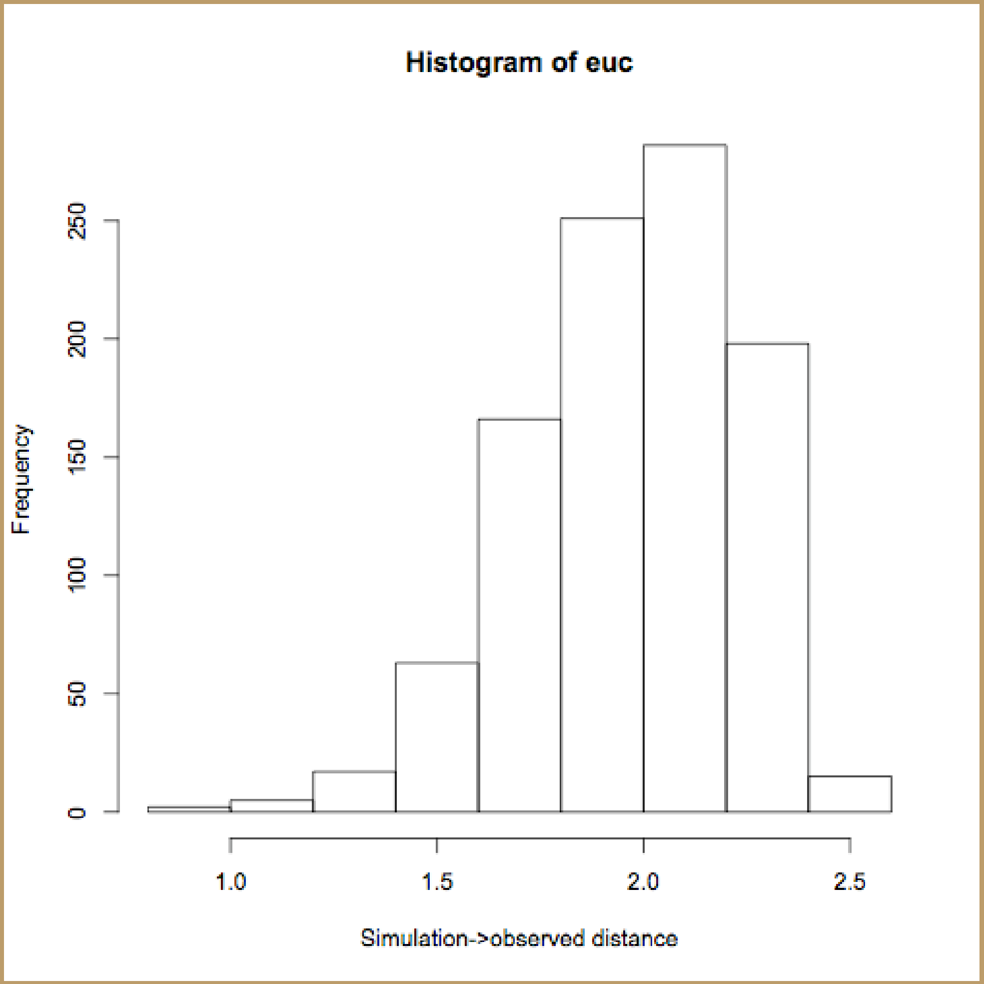
Delta:

Keep 1% of simulations

Type: 1.25

Press return to see next plot

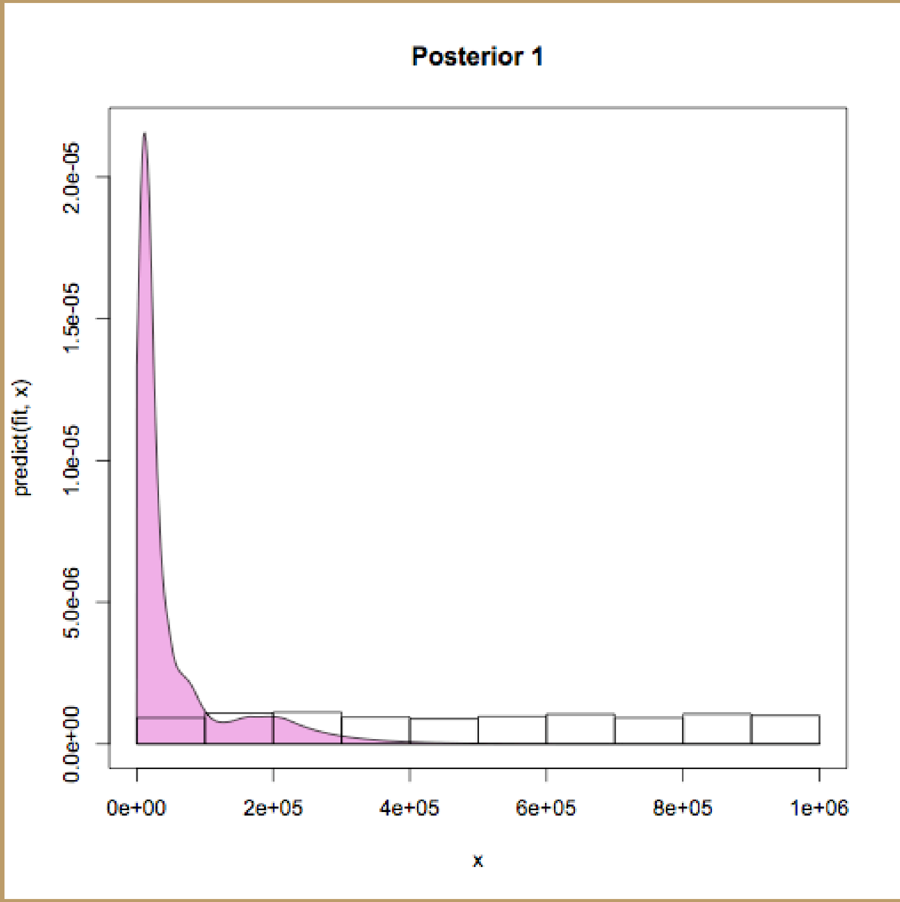
(don't close histogram window)



Using R to estimate posteriors

Your posterior for Ne

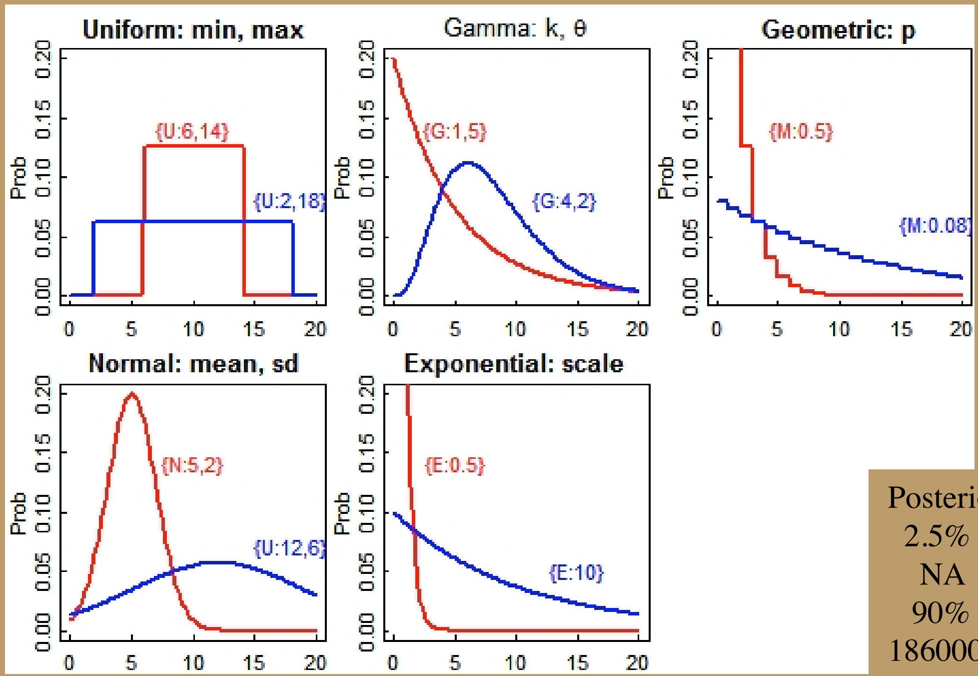
Let's run it again
{U:10,400000)

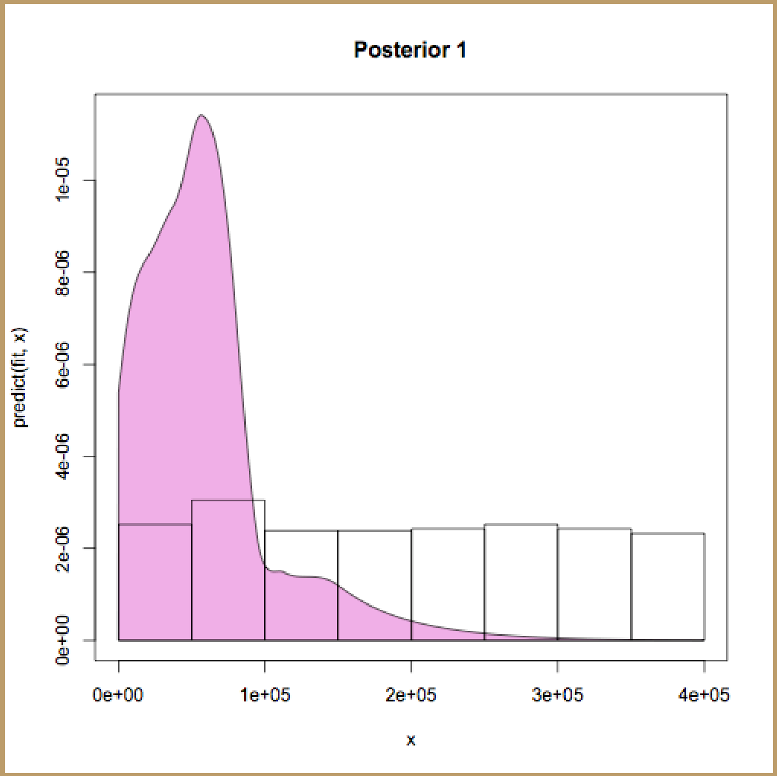


Using R to estimate posteriors

A closer look at your posterior for Ne

Now you need to choose a distribution





Posterior 1

2.5%	5%	10%	20%	30%	40%	50%	60%	70%	80%
NA	1300	4040	8640	13000	17600	23200	31800	49200	87400

90%	95%	97.5%
186000	239000	288000

{N:23000,7820}

Your final BayeSSC run for the null model

In your input file, replace your prior with your posterior and rerun BayeSSC

```
//Parameters for the coalescence simulation program : simcoal.exe
1 population with ancient DNA
//Population effective sizes (number of genes)
{N:23000,7820}
//Samples sizes:
4 sample groups
6 0 0 0
39 0 0 1
11 12500 0 2
3 12500 0 3
//Growth rates : negative growth implies population expansion
0
Number of migration matrices
0
historical event: time, source, sink, migrants, new deme size, new growth rate, migration matrix index
0 historical event
Mutation rate per generation for the whole sequence
0.0007872
Number of loci
615
//data type either DNA, RFLP, or MICROSAT : If DNA, we need a second term for the transition bias
DNA 0.792
//Gamma parameter (if 0: even mutation rates, if >0 :shape parameter of the Gamma distribution
0
```

Calculating model likelihood in R

Open R again

```
> SSC.Like("/Users/jessica/Tutorial/Tutorialnullposterior/Tutorialnullpost_stat.csv")
```

Calculating model likelihood in R

Open R again

```
> SSC.Like("/Users/jessica/Tutorial/Tutorialnullposterior/Tutorialnullpost_stat.csv")
 [1] "GROUP.0"        "Haptypes"       "PrivHaps"
 [4] "SegSites"       "PairDiffs"      "HapDiver"
 [7] "NucltdDiv"      "TajimasD"       "MismatDist"
[10] "X0.VS.1"        "PrivTo0"        "PrivTo1"
[13] "PairDiffs.1"    "MeanDiv.Hs.bar." "PoolDiv.Ht."
[16] "Fst"            "X0.VS.2"        "PrivTo0.1"
[19] "PrivTo2"        "PairDiffs.2"    "MeanDiv.Hs.bar..1"
[22] "PoolDiv.Ht..1"  "Fst.1"          "X0.VS.3"
[25] "PrivTo0.2"      "PrivTo3"        "PairDiffs.3"
[28] "MeanDiv.Hs.bar..2" "PoolDiv.Ht..2" "Fst.2"
[31] "GROUP.1"        "Haptypes.1"     "PrivHaps.1"
[34] "SegSites.1"     "PairDiffs.4"    "HapDiver.1"
[37] "NucltdDiv.1"    "TajimasD.1"     "MismatDist.1"
[40] "X1.VS.2"        "PrivTo1.1"      "PrivTo2.1"
[43] "PairDiffs.5"    "MeanDiv.Hs.bar..3" "PoolDiv.Ht..3"
[46] "Fst.3"          "X1.VS.3"        "PrivTo1.2"
[49] "PrivTo3.1"      "PairDiffs.6"    "MeanDiv.Hs.bar..4"
[52] "PoolDiv.Ht..4"  "Fst.4"          "GROUP.2"
[55] "Haptypes.2"     "PrivHaps.2"     "SegSites.2"
[58] "PairDiffs.7"    "HapDiver.2"     "NucltdDiv.2"
[61] "TajimasD.2"     "MismatDist.2"   "X2.VS.3"
[64] "PrivTo2.2"      "PrivTo3.2"      "PairDiffs.8"
[67] "MeanDiv.Hs.bar..5" "PoolDiv.Ht..5" "Fst.5"
[70] "GROUP.3"        "Haptypes.3"     "PrivHaps.3"
[73] "SegSites.3"     "PairDiffs.9"    "HapDiver.3"
[76] "NucltdDiv.3"    "TajimasD.3"     "MismatDist.3"
[79] "COMBINED"       "Haptypes.4"     "PrivHaps.4"
[82] "SegSites.4"     "PairDiffs.10"   "HapDiver.4"
[85] "NucltdDiv.4"    "TajimasD.4"     "MismatDist.4"
[88] "PRIORS"         "Deme.Size.0"
```

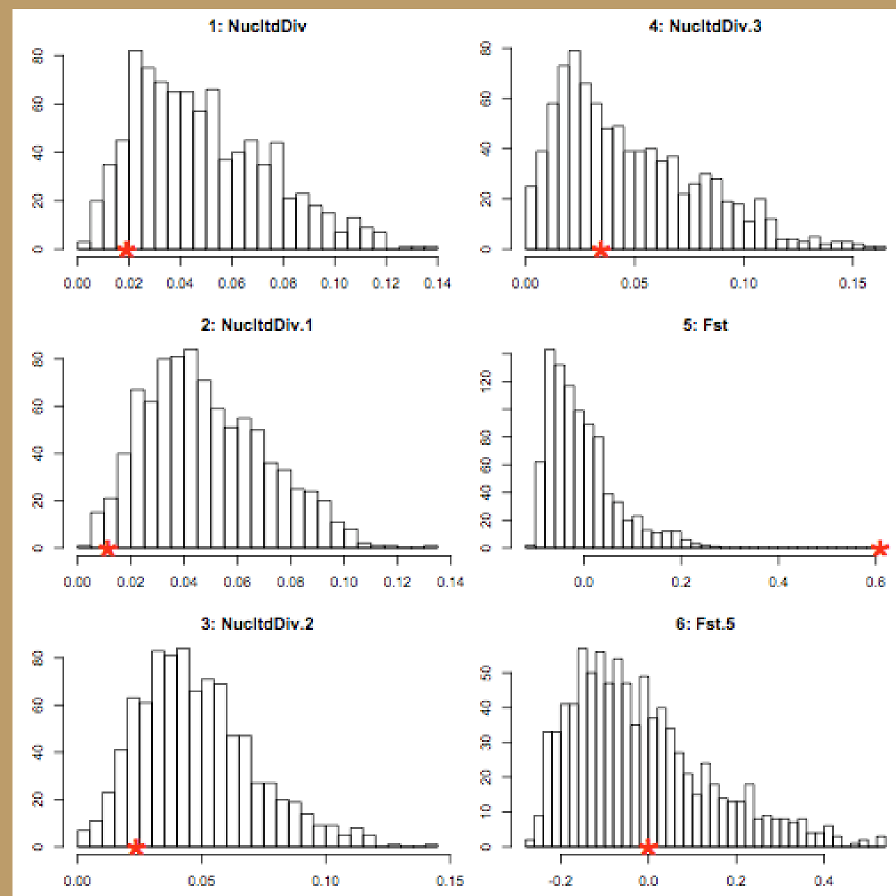
Calculating model likelihood in R

Open R again

```
> SSC.Like("/Users/jessica/Tutorial/Tutorialnullposterior/Tutorialnullpost_stat.csv")
 [1] "GROUP.0"        "Haptypes"        "PrivHaps"
 [4] "SegSites"        "PairDiffs"       "HapDiver"
 [7] "NucltdDiv"       "TajimasD"        "MismatDist"
[10] "X0.VS.1"         "PrivTo0"         "PrivTo1"
[13] "PairDiffs.1"     "MeanDiv.Hs.bar."   "PoolDiv.Ht."
[16] "Fst"             "X0.VS.2"         "PrivTo0.1"
[19] "PrivTo2"         "PairDiffs.2"     "MeanDiv.Hs.bar..1"
[22] "PoolDiv.Ht..1"    "Fst.1"           "X0.VS.3"
[25] "PrivTo0.2"       "PrivTo3"         "PairDiffs.3"
[28] "MeanDiv.Hs.bar..2" "PoolDiv.Ht..2"    "Fst.2"
[31] "GROUP.1"         "Haptypes.1"      "PrivHaps.1"
[34] "SegSites.1"      "PairDiffs.4"     "HapDiver.1"
[37] "NucltdDiv.1"     "TajimasD.1"      "MismatDist.1"
[40] "X1.VS.2"         "PrivTo1.1"       "PrivTo2.1"

.....

Which column/s (eg 4,23,27)? 7,37,60,76,16,69
Observed values: 0.01931,0.0116,0.024,0.035,0.61,0.00063
[1] 0
```

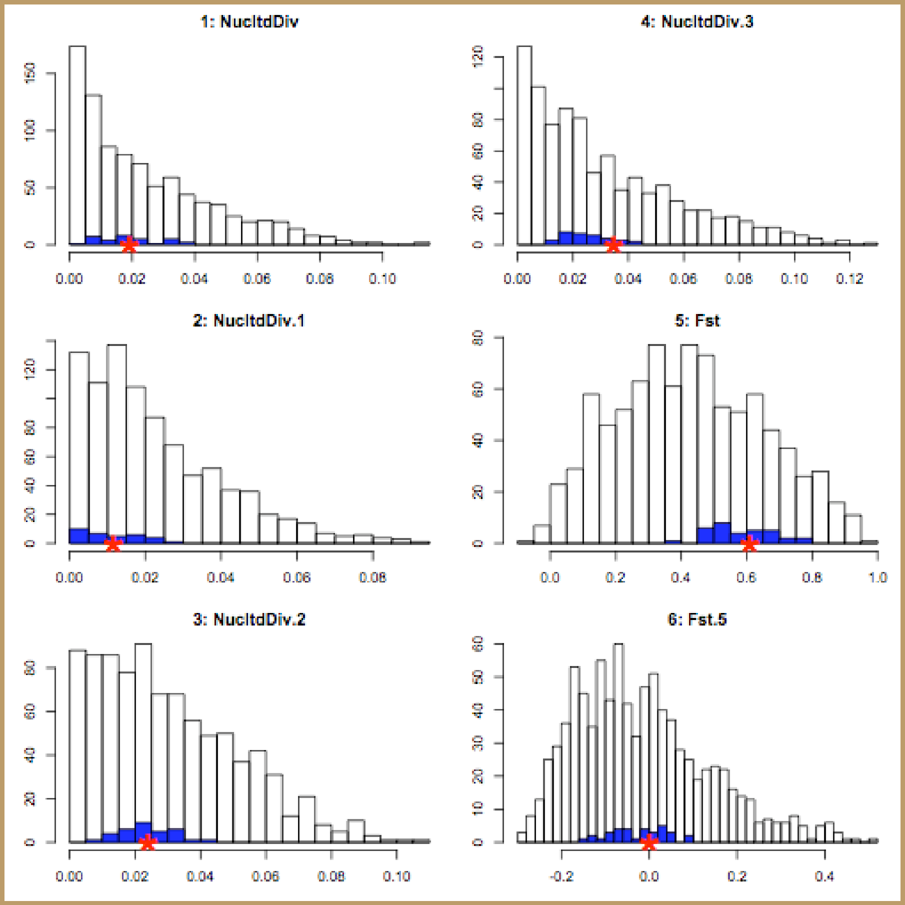


$$AIC = 2K - \log(\text{likelihood})$$
$$= 2 - \log(\text{likelihood})$$

Likelihood of Hypothesis 1

Open R again

Which column/s (eg 4,23,27)? 7,37,60,76,16,69
Observed values: 0.01931,0.0116,0.024,0.035,0.61,0.00063
[1] 0.03699552
> 2-log(0.036995)
[1] 5.296973



Word.