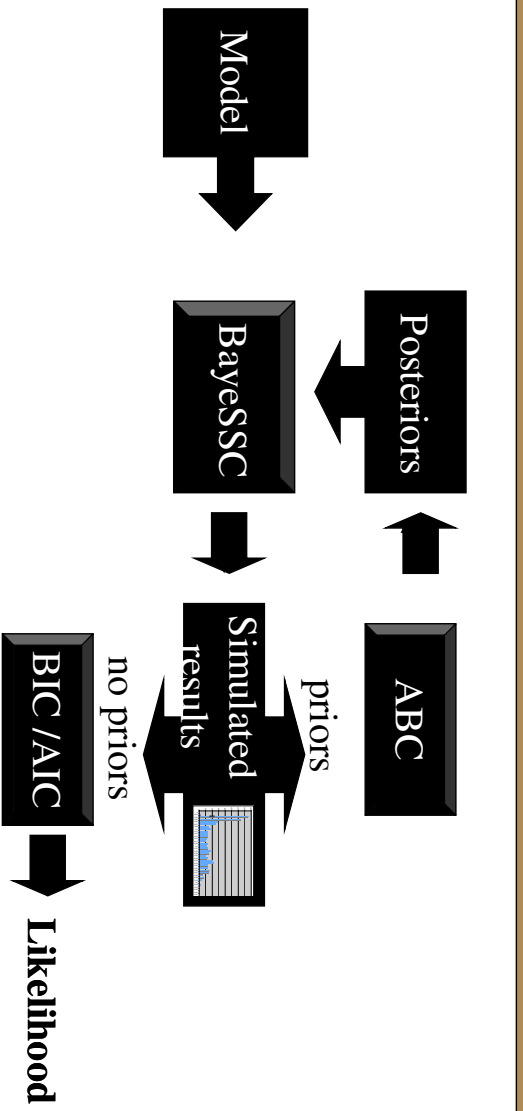


Tutorial BayeSSC



ppt originally designed and written by
Jessica Metcalf

How likely is each model?



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
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
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 simulation
1 population with ancient DNA
//Population effective size, 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 implies
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 MIG
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
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

```
//Parameters for the coalescence simulation
1 population with ancient DNA
//Population effective size, 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 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
```

$N(t)=N(0)e^{rt}$
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 coalescence simulation
1 population with ancient DNA
//Population effective size, 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 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
1 population with ancient DNA
(U:10,1000000)
//Population effective sizes (namb
(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
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
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

```
//Parameters for the coalescence
1 population with ancient DNA
(U:10,1000000)
//Population effective sizes (namb
(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
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
DNA 0.792
//Gamma parameter (if 0: even mutation rates, if >0: shape parameter of the Gamma distribution
0
```

1. The time (in generations) when the event occurred
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. Remember, 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 populat
//Popul
(U:1,0)
//Sample
4 sample
6 0 0 0
39 0 1 1
11 1 2500
3 1 500
//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
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
```

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

Input File

```
//Parameters for the coalescence simulation program : simcoal.exe
1 population with ancient DNA
//Population effective sizes (number of genes)
(U:1,0,1000000)
//Samples sizes:
4 sample groups
6 0 0 0
39 0 0 1
11 1 2500 0 2
3 1 2500 0 3
//Growth rates: negative growth implies population expansion
0
Number of migration matrices
0
historical event
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
```

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

Input File

```
//Parameters for the coalescence simulation program : simcoal.exe
1 popu
//Popul
(U:1,1
//Sam
4 sam
6.0 0.7
39.0 1.1
11.1254
3.1254
//Gewe
0
//Number of loci:
DNA 0.33333

* 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 heterogeneous (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

//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 0.0
11.1254
3.1254
//Gewe
0
//Number of loci:
DNA 0.33333

these 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

//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
```


Using R to estimate posteriors

```
> reject("~/Users/jessica/Tutorial/tutorialnull500k/tutorialnull_stat.csv")
Loading required package: locfit
Loading required package: akima
Loading required package: lattice
locfit 1.5-4      2007-11-27
111 "GROUP.0"      "PrivHaps"
141 "SegSites"    "ParDfns"
171 "NuclDiv"     "TajimasD"
1101 "X0.VS.1"    "PrivTo"
1131 "ParDfns.1"  "MeanDiv.Hs.bar." "PoolDiv.H."
1161 "Fst"        "X0.VS.2"      "PrivTo1"
1191 "PrivTo2"    "ParDfns.2"    "MeanDiv.Hs.bar.1"
1221 "PoolDiv.H.1" "Fst.1"        "X0.VS.3"
1251 "PrivTo0.2"  "PrivTo3"      "ParDfns.3"
1281 "MeanDiv.Hs.bar.2" "PoolDiv.H.2" "Fst.2"
1311 "GROUP.1"    "HapHaps.1"    "PrivHaps.1"
1341 "SegSites.1" "ParDfns.4"    "HapDiver.1"
1371 "NuclDiv.1" "TajimasD.1"   "MismatDist.1"
1401 "X1.VS.2"    "PrivTo1.1"    "PrivTo2.1"
1431 "ParDfns.5"  "MeanDiv.Hs.bar.3" "PoolDiv.H.3"
1461 "Fst.3"      "X1.VS.3"      "PrivTo1.2"
1491 "PrivTo3.1"  "ParDfns.6"    "MeanDiv.Hs.bar.4"
1521 "PoolDiv.H.4" "Fst.4"        "GROUP.2"
1551 "HapHaps.2"  "PrivHaps.2"   "SegSites.2"
1581 "ParDfns.7"  "HapDiver.2"   "NuclDiv.2"
1611 "TajimasD.2" "MismatDist.2" "X2.VS.3"
1641 "PrivTo2.2"  "PrivTo3.2"    "ParDfns.8"
1671 "MeanDiv.Hs.bar.5" "PoolDiv.H.5" "Fst.5"
1701 "GROUP.3"    "HapHaps.3"    "PrivHaps.3"
1731 "SegSites.3" "ParDfns.9"    "HapDiver.3"
1761 "NuclDiv.3"  "TajimasD.3"   "MismatDist.3"
1791 "COMBINED"  "HapHaps.4"    "PrivHaps.4"
1821 "SegSites.4" "ParDfns.10"   "HapDiver.4"
1851 "NuclDiv.4"  "TajimasD.4"   "MismatDist.4"
1881 "PRIORS"     "Deme.Size.0"
```

Which column/s (eg 4,23,27)?

Using R to estimate posteriors

```
> reject("~/Users/jessica/Tutorial/tutorialnull500k/tutorialnull_stat.csv")
```

Loading required package: locfit

Loading required package: akima

Loading required package: lattice

locfit 1.5-4 2007-11-27

```
111 "GROUP.0"      "PrivHaps"
141 "SegSites"    "ParDfns"
171 "NuclDiv"     "TajimasD"
1101 "X0.VS.1"    "PrivTo"
1131 "ParDfns.1"  "MeanDiv.Hs.bar." "PoolDiv.H."
1161 "Fst"        "X0.VS.2"      "PrivTo1"
1191 "PrivTo2"    "ParDfns.2"    "MeanDiv.Hs.bar.1"
1221 "PoolDiv.H.1" "Fst.1"        "X0.VS.3"
1251 "PrivTo0.2"  "PrivTo3"      "ParDfns.3"
1281 "MeanDiv.Hs.bar.2" "PoolDiv.H.2" "Fst.2"
1311 "GROUP.1"    "HapHaps.1"    "PrivHaps.1"
1341 "SegSites.1" "ParDfns.4"    "HapDiver.1"
1371 "NuclDiv.1" "TajimasD.1"   "MismatDist.1"
1401 "X1.VS.2"    "PrivTo1.1"    "PrivTo2.1"
1431 "ParDfns.5"  "MeanDiv.Hs.bar.3" "PoolDiv.H.3"
1461 "Fst.3"      "X1.VS.3"      "PrivTo1.2"
1491 "PrivTo3.1"  "ParDfns.6"    "MeanDiv.Hs.bar.4"
1521 "PoolDiv.H.4" "Fst.4"        "GROUP.2"
1551 "HapHaps.2"  "PrivHaps.2"   "SegSites.2"
1581 "ParDfns.7"  "HapDiver.2"   "NuclDiv.2"
1611 "TajimasD.2" "MismatDist.2" "X2.VS.3"
1641 "PrivTo2.2"  "PrivTo3.2"    "ParDfns.8"
1671 "MeanDiv.Hs.bar.5" "PoolDiv.H.5" "Fst.5"
1701 "GROUP.3"    "HapHaps.3"    "PrivHaps.3"
1731 "SegSites.3" "ParDfns.9"    "HapDiver.3"
1761 "NuclDiv.3"  "TajimasD.3"   "MismatDist.3"
1791 "COMBINED"  "HapHaps.4"    "PrivHaps.4"
1821 "SegSites.4" "ParDfns.10"   "HapDiver.4"
1851 "NuclDiv.4"  "TajimasD.4"   "MismatDist.4"
1881 "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/tutorialnull500k/tutorialnull_stat.csv")
```

```
Loading required package: locfit
```

```
Loading required package: akima
```

```
Loading required package: lattice
```

```
locfit 1.5-4      2007-11-27
```

```

111 "GROUP.0"      "PrivHaps"
141 "SegSites"    "ParDifs"
171 "NuclDiv"     "TajimasD"
1101 "X0.VS.1"    "PrivTo"
1131 "ParDifs.1"  "MeanDiv.Hs.bar." "PoolDiv.Ht."
1161 "Fst"        "X0.VS.2"      "PrivTo.1"
1191 "PrivTo2"   "ParDifs.2"    "MeanDiv.Hs.bar.1"
1221 "PoolDiv.H.1" "Fst.1"       "X0.VS.3"
1251 "PrivTo.2"  "PrivTo.3"     "ParDifs.3"
1281 "MeanDiv.Hs.bar.2" "PoolDiv.Ht.2" "Fst.2"
1311 "GROUP.1"   "Hapypes.1"    "PrivHaps.1"
1341 "SegSites.1" "ParDifs.4"    "HapDiver.1"
1371 "NuclDiv.1" "TajimasD.1"   "MismatDist.1"
1401 "X1.VS.2"   "PrivTo.1.1"   "PrivTo.1"
1431 "ParDifs.5" "MeanDiv.Hs.bar.3" "PoolDiv.Ht.3"
1461 "Fst.3"     "X1.VS.3"      "PrivTo.1.2"
1491 "PrivTo.3.1" "ParDifs.6"    "MeanDiv.Hs.bar.4"
1521 "PoolDiv.Ht.4" "Fst.4"       "GROUP.2"
1551 "Hapypes.2"  "PrivHaps.2"   "SegSites.2"
1581 "ParDifs.7"  "HapDiver.2"   "NuclDiv.2"
1611 "TajimasD.2" "MismatDist.2" "X2.VS.3"
1641 "PrivTo.2.2" "PrivTo.3.2"   "ParDifs.8"
1671 "MeanDiv.Hs.bar.5" "PoolDiv.Ht.5" "Fst.5"
1701 "GROUP.3"   "Hapypes.3"    "PrivHaps.3"
1731 "SegSites.3" "ParDifs.9"    "HapDiver.3"
1761 "NuclDiv.3"  "TajimasD.3"   "MismatDist.3"
1791 "COMBINED"  "Hapypes.4"    "PrivHaps.4"
1821 "SegSites.4" "ParDifs.10"   "HapDiver.4"
1851 "NuclDiv.4"  "TajimasD.4"   "MismatDist.4"
1881 "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
```

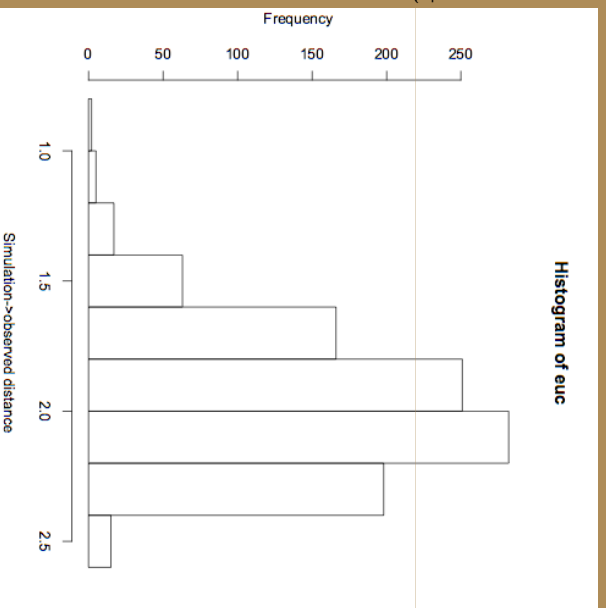
Statistic	csv column	Observed value
NuclDiv0	7	0.01931
NuclDiv1	37	0.0116
NuclDiv2	60	0.024
NuclDiv3	76	0.035
Fst.0v1	16	0.61
Fst.0v2	69	0.00063

Using R to estimate posteriors

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

Delta:

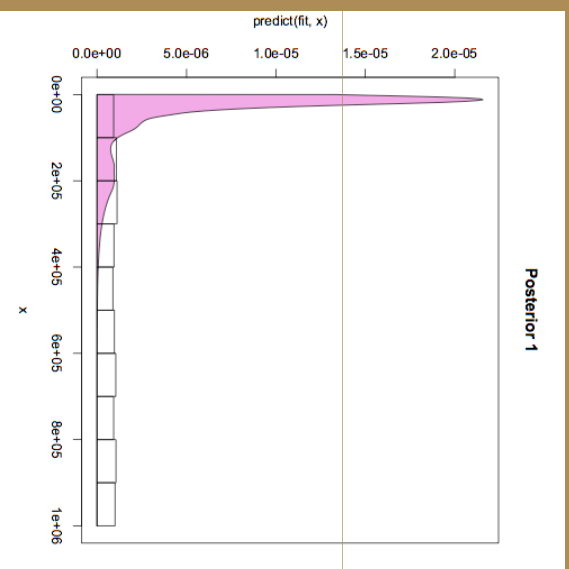
Keep 1% of simulations
 Type: 1.25
 Press return to see next plot
 (don't close histogram window)



Using R to estimate posteriors

Your posterior for Ne

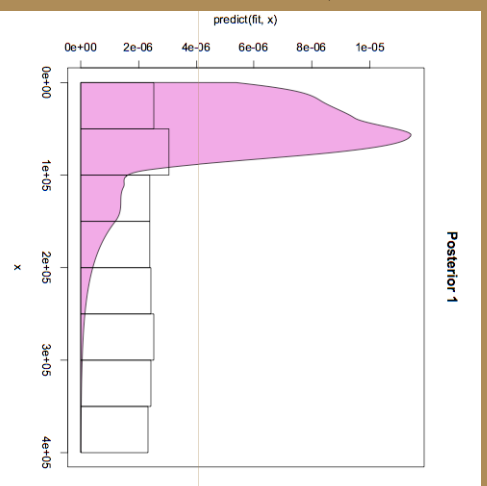
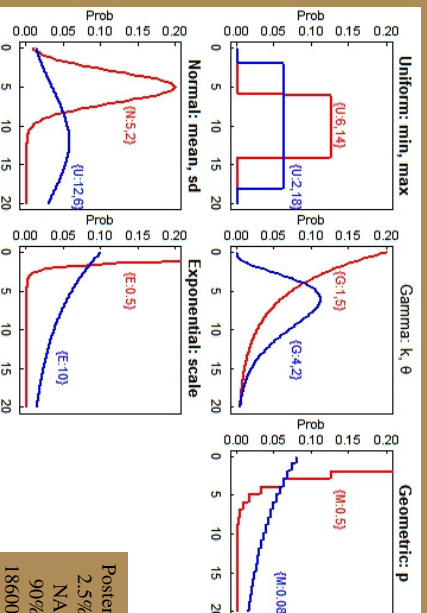
Let's run it again
 {U:10,400000}



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%
90%	95%
97.5%	NA
1300	4040
13000	8640
17600	13000
23200	17600
31800	23200
49200	31800
87400	49200
186000	87400
239000	186000
288000	239000

{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
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/Tutorialmultposterior/Tutorialmult_post_statcsv")
[1] "GROUP.0"      "Happypes"      "PrivHaps"
[4] "SegSites"     "PairDifs"     "HapDiver"
[7] "Nucldiv"      "TajimasD"     "MisnatDist"
[10] "X0.VS.1"      "PrivTo0"      "PrivTo1"
[13] "PairDifs.1"   "MeanDiv.Hs.bar." "PoolDiv.Ht."
[16] "Fst"          "X0.VS.2"      "PrivTo0.1"
[19] "PrivTo2"      "PairDifs.2"   "MeanDiv.Hs.bar.1"
[22] "PoolDiv.Ht.1" "Fst.1"        "X0.VS.3"
[25] "PrivTo0.2"   "PrivTo3"      "PairDifs.3"
[28] "MeanDiv.Hs.bar.2" "PoolDiv.Ht.2" "Fst.2"
[31] "GROUP.1"     "Happypes.1"  "PrivHaps.1"
[34] "SegSites.1"  "PairDifs.4"  "HapDiver.1"
[37] "Nucldiv.1"   "TajimasD.1"  "MisnatDist.1"
[40] "X1.VS.2"     "PrivTo1.1"   "PrivTo2.1"
[43] "PairDifs.5" "MeanDiv.Hs.bar.3" "PoolDiv.Ht.3"
[46] "Fst.3"      "X1.VS.3"     "PrivTo1.2"
[49] "PrivTo3.1"   "PairDifs.6"  "MeanDiv.Hs.bar.4"
[52] "PoolDiv.Ht.4" "Fst.4"       "GROUP.2"
[55] "Happypes.2" "PrivHaps.2"  "SegSites.2"
[58] "PairDifs.7"  "HapDiver.2"  "Nucldiv.2"
[61] "TajimasD.2"  "MisnatDist.2" "X2.VS.3"
[64] "PrivTo2.2"  "PrivTo3.2"   "PairDifs.8"
[67] "MeanDiv.Hs.bar.5" "PoolDiv.Ht.5" "Fst.5"
[70] "GROUP.3"    "Happypes.3"  "PrivHaps.3"
[73] "SegSites.3" "PairDifs.9"  "HapDiver.3"
[76] "Nucldiv.3"  "TajimasD.3"  "MisnatDist.3"
[79] "COMBINED"  "Happypes.4"  "PrivHaps.4"
[82] "SegSites.4" "PairDifs.10" "HapDiver.4"
[85] "Nucldiv.4"  "TajimasD.4"  "MisnatDist.4"
[88] "PRIORS"     "Denne.Size.0"
```

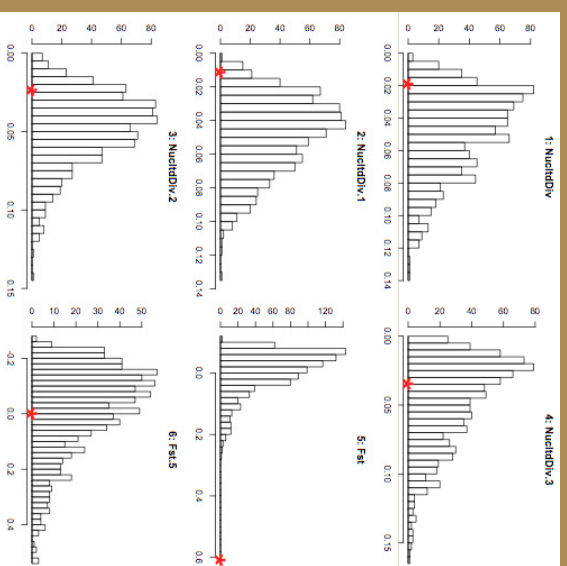
Calculating model likelihood in R

Open R again

```
> SSC.Like("Users/jessica/Tutorial/Tutorialmultposterior/Tutorialmult_post_statcsv")
[1] "GROUP.0"      "Happypes"      "PrivHaps"
[4] "SegSites"     "PairDifs"     "HapDiver"
[7] "Nucldiv"      "TajimasD"     "MisnatDist"
[10] "X0.VS.1"      "PrivTo0"      "PrivTo1"
[13] "PairDifs.1"   "MeanDiv.Hs.bar." "PoolDiv.Ht."
[16] "Fst"          "X0.VS.2"      "PrivTo0.1"
[19] "PrivTo2"      "PairDifs.2"   "MeanDiv.Hs.bar.1"
[22] "PoolDiv.Ht.1" "Fst.1"        "X0.VS.3"
[25] "PrivTo0.2"   "PrivTo3"      "PairDifs.3"
[28] "MeanDiv.Hs.bar.2" "PoolDiv.Ht.2" "Fst.2"
[31] "GROUP.1"     "Happypes.1"  "PrivHaps.1"
[34] "SegSites.1"  "PairDifs.4"  "HapDiver.1"
[37] "Nucldiv.1"   "TajimasD.1"  "MisnatDist.1"
[40] "X1.VS.2"     "PrivTo1.1"   "PrivTo2.1"
.....
```

Which columns/ (eg 4,2,3,27)? 7.37,60,76,16,69
 Observed values: 0.01931,0.0116,0.024,0.035,0.61,0.00063
 [1] 0

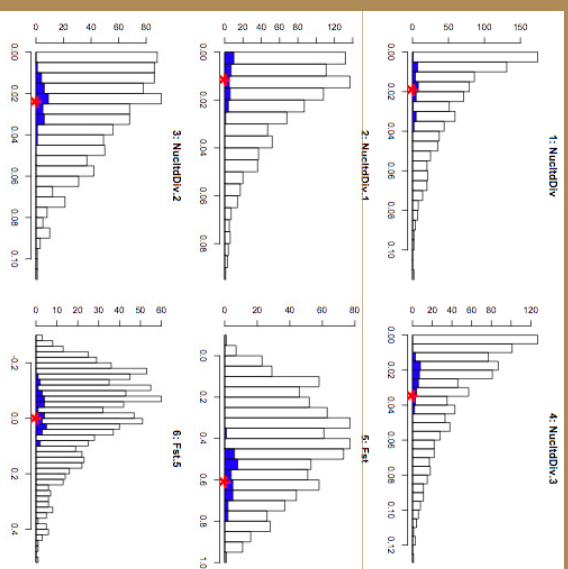
$$AIC = 2K - \log(\text{likelihood}) = 2 - \log(\text{likelihood})$$



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
```



Alternative Likelihoods

In the very strictest terms, calculating AIC this way is not correct if distributions are strongly skewed away from the fit.

Rather than putting “fit distributions” into the parameter file, put in the MLE, then calculate the AIC

```
reject (“mystuff_stat.csv”) -> z
mle(z)
```

(PUT MLE INPUT INTO PAR FILE AND RERUN)

```
aic.ssc (“mystuff_mle_stat.csv”)
```

Word.